



**HAL**  
open science

# Etude structurale et enzymatique du complexe entre l'ARNtSec et la Sélénocystéine Synthase SelA

Stéphane Sandeau-Gruber

► **To cite this version:**

Stéphane Sandeau-Gruber. Etude structurale et enzymatique du complexe entre l'ARNtSec et la Sélénocystéine Synthase SelA. Biochimie, Biologie Moléculaire. Université Paris Sud - Paris XI, 2008. Français. NNT : 2008PA112347 . tel-00657412

**HAL Id: tel-00657412**

**<https://theses.hal.science/tel-00657412>**

Submitted on 6 Jan 2012

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

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

UNIVERSITÉ PARIS XI

École doctorale de Chimie Paris Sud

# Étude structurale et enzymatique du complexe entre l'ARN<sub>t</sub><sup>Sec</sup> et la Sélénocystéine Synthase SelA.

## THÈSE

soutenue le 10 Juillet 2008

pour l'obtention du grade de

DOCTEUR EN SCIENCE DE L'UNIVERSITÉ PARIS XI

par

M GRUBER Stéphane  
époux SANDEAU-GRUBER

### Composition du jury :

<i>Président :</i>	Pr. THERISOD Michel	ICMMO (Orsay)
<i>Rapporteurs :</i>	Dr. KROL Alain	IBMC (Strasbourg)
	Dr. CRÉVISY Christophe	ENSCR (Rennes)
<i>Examineur :</i>	Dr. CHAVATTE Laurent	CGM (Gif-sur-Yvette)
<i>Directeur de thèse :</i>	Dr. FOURMY Dominique	ICSN (Gif-sur-Yvette)
<i>Co-encadrant de thèse :</i>	Dr. YOSHIZAWA Satoko	ICSN (Gif-sur-Yvette)



## Remerciements

Aux sombres héros de l'amer  
Qui ont su traverser les océans du vide.

---

*Veillez rendre l'âme (à qui elle appartient)*  
**Noir Désir (1989)**

J'ai eu beau essayer, je crois que j'ai toujours eu du mal à imaginer le jour où je soutiendrai ma thèse. Bien sûr, quand j'assistais aux soutenances de collègues, je pensais bien qu'un jour mon tour viendrait. Bien sûr, à chaque fois que je parcourais un manuscrit, je commençais par lire les remerciements, pour voir ce que le nouveau docteur avait retiré comme expériences humaines de ce long cheminement, en essayant d'imaginer ce que, moi, j'écrirai. Et, aujourd'hui, j'y suis.

Ma soutenance, je ne l'avais pas imaginée, ni même rêvée. Elle s'est pourtant déroulée le mieux que j'aurais pu imaginer. J'aimerais remercier mon jury pour avoir favorisé ce jour-là l'ambiance sereine et amicale que je garderais en mémoire.

Mes remerciements vont ainsi à Michel Therisod pour avoir accepté de présider ce jury et de son intérêt pour mon travail alors que celui-ci ne relevait pas directement de sa spécialité, à Alain Krol pour ses remarques constructives et pointues sur le manuscrit et lors de ma soutenance, à Christophe Crévisy pour avoir accepté d'être également mon rapporteur et ainsi de m'accompagner une fois de plus dans une grande étape de mon cursus scientifique, et enfin à Laurent Chavatte dont la sympathie m'a suivie de notre première rencontre, à nos échanges scientifiques sur nos travaux, jusqu'à cette soutenance et ensuite.

Je remercie Dominique Fourmy et Satoko Yoshizawa pour m'avoir permis d'entrer dans l'univers fabuleux de la sélénocystéine, dont ma thèse n'a révélé à mes yeux qu'une petite portion et dont j'ai hâte d'en découvrir d'avantage. Merci de m'avoir accepté dans votre thématique et de m'avoir formé aux techniques de biologie moléculaire et de biochimie qui m'ont permis d'aboutir à ce titre de Docteur. Merci enfin pour vos encouragements lorsque les chemins que nous prenions semblaient sans issue.

Je tiens à remercier Jean-Yves Lallemand pour son soutien durant ce doctorat au sein de l'ICSN, et en particulier lors de ma décision de changer de thématique.

Un grand merci à Eric Guittet pour m'avoir accueilli dans son équipe et pour m'avoir permis de terminer ce doctorat dans les meilleures conditions possibles.

L'aventure scientifique que représente une thèse n'en est pas moins une aventure humaine. Celle-ci fut agréablement agrémentée par la convivialité qui règne au bâtiment 23B, imputable à tous mes acolytes du labo : à mes compagnons de rédaction de manuscrits, Dirk-EcoloPCmanEnSpartiate et Olivier-ToujoursLeMotPourRire-EtD'abordSiLaRMNne

MarchePasJeFeraisUneChouetteCarrièreEnInformatique(MêmePasMal), à mon cher co-voitureur Aurélien qui est largement plus fiable que ma voiture (et beaucoup plus sympa), à mes collègues de sous-équipe ARN : Nicolas<sup>3</sup>, Clotilde, Cyril, Bewein ; et à tous ceux qui sont passés sous la bannière de l'équipe 64 (Benoît, Stéphanie, Matthieu, Sophie, Vanessa, FX, Diane, la chance...) et ceux qui y sont titulaire et qui font de l'équipe ce qu'elle est (Jean-Louis, Ewen, Nicolas, Carine, en plus de ceux déjà cités).

Une mention spéciale à Alda dont les compétences, la disponibilité et la bonne humeur inébranlable méritent qu'elle serve d'exemple dans le Robert au terme "perle".

Bien sûr, je souhaite remercier toutes les personnes extra-64 qui ont peuplé ma vie de labo ces 3 dernières années (et demie), et j'aimerais dire un mot personnel à chacune d'elle. Mais par soucis de concision et pour ne pas devenir ennuyeux, je vais faire court ; celles qui sont citées ici le méritent, par leur rôle dans l'égaiement de mes journées (ou soirées) : Pierre, Laetitia, Edith et Jean-Denis, Damjana, Christina, Jean-Marie, Louise, Annie, Eric Jacquet, Jean-François Gallard, Guillaume Loire, et les autres...

Le bâtiment 23B n'est pas le lieu du début de ma vie de thésard : je tiens à remercier tous ceux qui ont fait un peu plus que croiser mon chemin au bâtiment 27.

En particulier, il me tient à coeur de réserver une pensée aux membres de mon premier laboratoire de thèse, au sein de l'équipe 63, sans lesquels je ne serais pas venu débiter une thèse de chimie en région parisienne, et n'aurais jamais pu, par la suite, m'ouvrir à la bio-mol et à la biochimie, en changeant de thématique de thèse. Même si je n'ai pas pu approfondir ce sujet, j'ai eu de l'intérêt pour la photochimie des nucléosides.

Merci également à tous les thésards de l'Institut qui ont été suffisamment indulgents pour me supporter lorsque je présidais le CEI 2005, surtout ceux du CEI 2004 pour leur accueil sympathique et leurs soirées jeux qui permettaient de s'évader après une dure journée de labeur.

Au sein de ce groupe se cachaient des êtres merveilleux, qui m'ont accompagné tout au long de mon périple de thésard, qui ont partagé la sueur et la poussière de plusieurs éditions des 24h du Mans Rollers, et qui continue de m'accompagner. Donc, un gros poutou à Pierre, et un énorme à Guillaume qui a poussé l'abnégation au point de me faire le plaisir de relire (et corriger) mon manuscrit.

Un grand merci également à ceux qui m'ont entouré de gentillesse depuis très longtemps ou depuis peu : Matthieu, Ben & Nat.

Enfin, un énorme merci à ma famille, qui m'a conduit vers ce que je suis devenu. Et surtout, pour Anne-Caroline, les mots me manquent pour décrire ta part dans ce travail : sans toi, je n'y serais jamais arrivé. À toi et Gabriel, qui n'est pas encore tout à fait là, bien que tu occupes une place de plus en plus importante dans ce monde et dans nos vies. Cette thèse est un peu ta jumelle...

*Coordonnées permanentes : [stephane.sandean-gruber@polytechnique.edu](mailto:stephane.sandean-gruber@polytechnique.edu)*

*À ma famille, à l'ange et demi qui m'ont donné la volonté d'y parvenir...*



# Glossaire

Termes et acronymes utilisés dans ce manuscrit.

<b>aaRS</b> : aminoacyl-ARN <sub>t</sub> -synthétase	<b>LUCA(S)</b> : Last Universal Common Ancestor (State)
<b>ADN</b> : Acide désoxyribonucléique	<b>MMA</b> : Monométhylamine
<b>AdT</b> : Amidotransférase	<b>MPD</b> : 2-méthyl-2,4-pentanediol
<b>AJR</b> : Apports Journaliers Recommandés	<b>Msr</b> : Méthionine-sulfoxyde réductase
<b>ARN</b> : Acide ribonucléique	<b>NMIA</b> : Anhydride N-méthylisatoïque
<b>ARN<sub>m</sub></b> : ARN messager	<b>ORF</b> : Open reading frame
<b>ARN<sub>t</sub></b> : ARN de transfert	<b>PAGE</b> : Électrophorèse sur gel de polyacrylamide
<b>ATP</b> : Adénosine triphosphate	<b>PCR</b> : Polymerase chain reaction
<b>CCM</b> : Chromatographie sur couche mince	<b>PDB</b> : Protein Data Bank
<b>CPM</b> : Coups par minute	<b>PEG</b> : Polyéthylène glycol
<b>Da</b> : Dalton (kDa : kiloDalton)	<b>pI</b> : point isoélectrique
<b>DL50</b> : Dose létale pour 50% d'une population.	<b>PLP</b> : Pyridoxal-5'-phosphate
<b>DMA</b> : Diméthylamine	<b>RF</b> : Facteur de terminaison
<b>DTT</b> : Dithiothreitol : 1,4-dimercaptobutane-2,3-diol	<b>RMN</b> : Résonance Magnétique Nucléaire
<b>EF-G</b> : Facteur de translocation	<b>RRF</b> : Facteur de recyclage du ribosome
<b>EF-Tu</b> : Facteur d'élongation Tu	<b>SCL</b> : Sélénocystéine lyase
<b>ESRF</b> : European synchrotron radiation facility	<b>SDS</b> : Dodécylsulfate de sodium
<b>Fdh</b> : Formate déshydrogénase	<b>Sec ou U</b> : Sélénocystéine
<b>FPLC</b> : Fast protein liquid chromatography	<b>SelA</b> : Sélénocystéine synthase
<b>GPx</b> : Glutathion peroxydase	<b>Sep</b> : Phosphosérine
<b>HEPES</b> : Acide N-(2-hydroxyéthyl)-piperazine-N'-(2-éthanesulfonique)	<b>SepCysS</b> : Sep-ARN <sub>t</sub> <sup>Cys</sup> Cys-ARN <sub>t</sub> <sup>Cys</sup> synthase
<b>HPLC</b> : High pressure liquid chromatography	<b>SepSecS</b> : Sep-ARN <sub>t</sub> <sup>Sec</sup> Sec-ARN <sub>t</sub> <sup>Sec</sup> synthase
<b>IF</b> : Facteur d'initiation de la traduction	<b>SHAPE</b> : Selective 2'-Hydroxyl Acylation analyzed by Primer Extension
<b>ITC</b> : Calorimétrie par titration isotherme	<b>SIDA</b> : Syndrome d'immunodéficience acquise
	<b>SIM</b> : Single Injection Method (ITC)



**SPS** : Sélénophosphate synthétase

**TA** : Température ambiante

**TCA** : Acide trichloroacétique

**TMA** : Triméthylamine

**TRIS** : Trishydroxyméthylaminométhane

**TrxR** : Thiorédoxine réductase

**U.V.** : Ultraviolet

**VIH** : Virus de l'immunodéficience humaine

# Table des matières

---

---

## Partie I

---

---

### Introduction générale

1	Le sélénium . . . . .	18
2	Toxicité et déficience . . . . .	19
3	Les végétaux et le sélénium . . . . .	21
4	Le sélénium dans nos cellules . . . . .	23

### Chapitre 1

#### La sélénocystéine

1.1	Un acide aminé tardivement découvert . . . . .	34
	présentation et historique . . . . .	34
1.2	Rappels sur le code génétique et la traduction . . . . .	35
	1.2.1 La synthèse protéique . . . . .	35
	1.2.2 Les 20 acides aminés canoniques : Le Code génétique . . . . .	38
1.3	La sélénocystéine, exception du code génétique . . . . .	40
	1.3.1 Sa biosynthèse sur l'ARN <sub>t</sub> <sup>Sec</sup> . . . . .	41
	1.3.2 Son incorporation co-translationnelle assistée . . . . .	43
	1.3.3 Le cycle de l'acide aminé libre . . . . .	47

1.4	Autres acides aminés particuliers comparables . . . . .	49
1.4.1	L'asparagine et la glutamine . . . . .	49
1.4.2	La pyrrolysine, le 22 <sup>ème</sup> Acide Aminé . . . . .	51
1.4.3	La cystéine chez certaines Archaeas . . . . .	55
1.5	L'avantage catalytique de la sélénocystéine . . . . .	60
1.5.1	Propriétés chimiques de la sélénocystéine . . . . .	60
1.5.2	Réactions enzymatiques impliquant une sélénocystéine . . . . .	63
1.6	Implications évolutives . . . . .	67
1.6.1	Les sélénoprotéines Archaeas . . . . .	68
1.6.2	Une grande disparité des sélénoprotéomes eucaryotes . . . . .	69
1.6.3	Hypothèses sur les facteurs entraînant cette disparité . . . . .	73
1.6.4	Confirmation de ces tendances par l'étude des eubactéries . . . . .	74

**Chapitre 2**

**Les acteurs de la synthèse de la sélénocystéine sur son ARN<sub>t</sub>**

2.1	L'ARN <sub>t</sub> <sup>Sec</sup> . . . . .	81
2.1.1	Structures et déterminants . . . . .	81
2.1.2	L'importance de la modification des bases . . . . .	85
2.1.3	Évolution et lien entre les ARN <sub>t</sub> <sup>Sec</sup> . . . . .	86
2.2	Le sélénophosphate SePO <sub>3</sub> <sup>3-</sup> . . . . .	87
2.2.1	Présentation et historique . . . . .	87
2.2.2	Propriétés chimiques . . . . .	88
2.2.3	La sélénophosphate synthétase . . . . .	89
2.3	La sélénocystéine synthase Sela . . . . .	90
2.3.1	Présentation . . . . .	90
2.3.2	Dans la famille des enzymes PLP-dépendantes, je demande... . . . .	92
2.3.3	Mécanisme . . . . .	94
2.3.4	Stœchiométrie et discrimination chez Sela . . . . .	94
2.3.5	Quel lien évolutif avec les enzymes homologues ? . . . . .	98
2.3.6	Rares sont les structures décameriques . . . . .	106
2.4	Conclusion . . . . .	107

---

---

## Partie II

---

---

### Chapitre 3

#### Vers l'obtention de cristaux du complexe SelA / $\text{ARN}_t^{\text{Sec}}$

3.1	Choix des partenaires . . . . .	132
3.1.1	Les $\text{ARN}_t^{\text{Sec}}$ de <i>E. coli</i> et <i>M. thermoacetica</i> . . . . .	133
3.1.2	Caractéristiques des sélénocystéine synthases SelA . . . . .	134
3.2	Obtention des $\text{ARN}_t^{\text{Sec}}$ . . . . .	135
3.2.1	Au départ, <i>selC</i> ... . . . .	135
3.2.2	La transcription . . . . .	139
3.2.3	La charge en sérine . . . . .	140
3.2.4	... au final, le séryl- $\text{ARN}_t^{\text{Sec}}$ . . . . .	142
3.3	Obtention des sélénocystéine synthases . . . . .	144
3.3.1	Clonage et vecteur d'expression . . . . .	144
3.3.2	Production et récolte . . . . .	144
3.3.3	Purification des constructions protéiques . . . . .	145
3.4	Formation du complexe . . . . .	153
3.4.1	Mise en présence des partenaires . . . . .	153
3.4.2	Purification du complexe . . . . .	154
3.4.3	Essais de production directe du complexe . . . . .	155
3.5	Cristallogénèse . . . . .	156
3.5.1	Techniques employées . . . . .	157
3.5.2	Carnet de bord de la cristallogénèse . . . . .	160
3.5.3	Cristaux et reproductibilité . . . . .	169

### Chapitre 4

#### Etudes enzymatiques du couple SelA / $\text{ARN}_t^{\text{Sec}}$

4.1	Formation du complexe . . . . .	174
4.1.1	Observation directe des produits de réaction . . . . .	174

4.1.2	Synthèse chimique du sélénophosphate $\text{SePO}_3^{3-}$ . . . . .	176
4.2	Mesures de protection à la déacylation . . . . .	179
4.2.1	Technique employée . . . . .	180
4.2.2	Résultats obtenus . . . . .	181
4.3	Expériences d'empreintes chimiques . . . . .	182
4.3.1	La technique SHAPE . . . . .	183
4.3.2	Empreintes sur l'ARN <sub>t</sub> <sup>Sec</sup> libre et complexé . . . . .	184
4.3.3	Implications des résultats . . . . .	186
4.4	Cinétiques de la réaction de complexation . . . . .	188
4.4.1	Rappel de la problématique . . . . .	188
4.4.2	Techniques de mesure . . . . .	190
4.4.3	Mesures effectuées et résultats obtenus . . . . .	192
4.4.4	Modélisation du comportement enzymatique de SelA . . . . .	197

<b>Chapitre 5</b>
-------------------

<b>Discussion sur la biosynthèse de la sélénocystéine chez les eubactéries</b>
--

5.1	Portée des résultats expérimentaux obtenus . . . . .	206
5.1.1	Au niveau structural... . . . .	207
5.1.2	Analyse du fonctionnement enzymatique de SelA . . . . .	208
5.2	Étude structurale théorique sur SelA . . . . .	210
5.2.1	Calculs de prédictions des structures secondaires de SelA . . . . .	211
5.2.2	Comparaisons structurales avec les protéines homologues . . . . .	212
5.2.3	Vers un modèle structural de SelA ? . . . . .	219
5.3	Perspectives . . . . .	220

---

---

## Annexes

---

---

<b>Annexe A</b>	<b>Annexes bibliographiques</b>	<b>229</b>
A.1	Arbre phylogénique de procaryotes séquencés . . . . .	229
A.2	Locus et opérons des gènes <i>selABCD</i> . . . . .	229
A.3	Phylogénie de SelA chez les eubactéries . . . . .	232
A.4	Les séquences protéiques de ecSelA et de mtSelA . . . . .	232
A.5	Arbres phylogéniques des enzymes PLP-dépendantes . . . . .	234
A.6	Structures d'enzymes PLP-dépendantes . . . . .	235
A.7	Structures des protéines décameriques connues . . . . .	241
<b>Annexe B</b>	<b>Matériels et Méthodes</b>	<b>245</b>
B.1	Fiches techniques . . . . .	245
B.1.1	Plasmide utilisé pour le clonage de <i>selC</i> . . . . .	245
B.1.2	Plasmide utilisé pour le clonage de SelA . . . . .	246
B.2	Protocoles expérimentaux . . . . .	247
B.2.1	PCR de multiplication de la matrice de transcription . . . . .	247
B.2.2	Transcription de <i>selC</i> en $\text{ARN}_t^{\text{Sec}}$ . . . . .	248
B.2.3	Charge en sérine de l' $\text{ARN}_t^{\text{Sec}}$ . . . . .	249
<b>Annexe C</b>	<b>La chimie du sélénium</b>	<b>251</b>
	<b>Références bibliographiques des Annexes</b>	<b>253</b>
	<hr/>	
	<b>Index</b>	<b>255</b>



# Première partie





# Introduction générale

## Sommaire

---

<b>1</b>	<b>Le sélénium . . . . .</b>	<b>18</b>
<b>2</b>	<b>Toxicité et déficience . . . . .</b>	<b>19</b>
<b>3</b>	<b>Les végétaux et le sélénium . . . . .</b>	<b>21</b>
<b>4</b>	<b>Le sélénium dans nos cellules . . . . .</b>	<b>23</b>

---

Cette thèse s'inscrit dans le large et passionnant thème de recherche sur le vivant qu'est : la compréhension des tenants et aboutissants de la synthèse protéique via la traduction ribosomale. Ce système complexe et global, qui permet la construction de la machinerie cellulaire définissant biologiquement et moléculairement ce que nous sommes, renferme encore beaucoup d'inconnu. Pour appréhender la philosophie d'un système, en étudier les exceptions, expliquer leur existence et leur fonctionnement s'est souvent révélé utile. L'acide aminé sélénocystéine (Sec / U), anomalie dans l'univers de la synthèse protéique, peut participer à la compréhension de l'origine de ce qui rassemble les êtres vivants de notre planète : ce « Frozen Accident » qu'est le Code Génétique tel que le décrit Francis Harry Compton Crick en 1968,<sup>[1.1]</sup> à l'époque de son élucidation. C'est pourquoi, depuis la découverte en 1987<sup>[1.2-1.4]</sup> de cet état d'anomalie, l'étude de la sélénocystéine n'a cessé de rassembler un nombre grandissant de scientifiques.

La sélénocystéine, que nous présenterons dans le premier chapitre de ce manuscrit, intéresse également grandement la médecine, depuis la découverte de certaines propriétés du sélénium, composant qui en fait l'originalité. En effet, le sélénium est non-seulement un oligo-élément vital, mais est fortement impliqué dans les mécanismes de protection cellulaire contre le stress oxydant. Ainsi les notions de dérèglement, vieillissement et mort cellulaire sont en partie dépendantes de ce composé. Au-delà du mythe de la potion de jouvence ou d'un anti-âge miracle, il s'agit bien d'un moyen mis en place par l'organisme pour mieux se défendre contre radicaux, dégradations du matériel cellulaire et survenue de lignées cellulaires malignes. Beaucoup d'espairs reposent sur la compréhension de l'action de substances comme le sélénium pour apporter l'aide nécessaire à l'organisme, lors d'une défaillance de sa part dans la prise en charge de cette protection, et prévenir la survenue de cancer, voire en faciliter une possible guérison.

Faisons donc connaissance avec le sélénium, cette substance mi-remède, mi-poison.

# 1 Le sélénium

Tout est poison, rien n'est poison :  
le poison, c'est la dose.

---

*Claude Bernard (1813–1878)*  
*citant Paracelse (1493–1541)*

## 16 (VIA)

Selenium		
K 2	Se <sub>34</sub>	Pt Fus. 221°C
L 8		
M 18		Pt Ebu. 685°C
N 6		
M = 78,96		
Etats d'Ox. -2+0+4+6		
Abondance : 0,05 ppm		

Le sélénium est un élément chimique de la famille des chalcogènes, au même titre que l'oxygène, le soufre ou le tellure. Sa similitude avec ce dernier est d'ailleurs à l'origine de son nom. En 1817, le chimiste suédois Jöns Jacob Berzelius découvrit cette nouvelle substance élémentaire dans un résidu de préparation de l'acide sulfurique. Comme cet élément avait les mêmes propriétés que le tellure, qui est nommé en rapport à la Terre (du Latin : *tellus*), le sélénium fut baptisé en hommage à la Lune (du Grec :  $\Sigma\epsilon\lambda\eta\nu\eta$  "selènè").

Sur la surface de la Terre, le sélénium est présent de façon extrêmement variable dans les sols selon que l'on se trouve dans une zones séléniprives (teneurs inférieures à 0,1µg/g) ou dans une zones sélénifères où sa teneur peut atteindre jusqu'à 1200µg/g.

Isotopes	<sup>74</sup> Se	<sup>76</sup> Se	<sup>77</sup> Se	<sup>78</sup> Se	<sup>80</sup> Se	<sup>82</sup> Se
Proportions	0,87%	9,36%	7,63%	23,78%	49,61%	8,73%

**Tableau 1** – Occurrence isotopique naturelle du sélénium.

Les différentes formes que prend le sélénium sont identiques à celles du soufre : ce sont les sélénures ( $\text{Se}_2^-$  :  $n_{\text{ox}} -2$ ), les différentes formes allotropiques du sélénium élémentaire ( $\text{Se}$  :  $n_{\text{ox}} +0$ ), les sélénites ( $\text{SeO}_3^{2-}$  :  $n_{\text{ox}} +4$ ) et les sélénates ( $\text{SeO}_4^{2-}$  :  $n_{\text{ox}} +6$ ).

Le sélénium élémentaire se retrouve sous plusieurs formes :

- le sélénium rouge amorphe, qui peut être obtenu par précipitation.
- le sélénium vitreux, masse amorphe brun-noir qui se forme par refroidissement brutal du sélénium liquide. Il est non-conducteur.
- le sélénium gris ou improprement appelé sélénium métallique, cristal du système trigonal, qui est la variété thermodynamiquement stable : au-dessus de 110°C, les autres variétés deviennent du sélénium cristallin. Il peut être obtenu par refroidissement lent du sélénium liquide. De densité 4,80 , il est utilisé pour ses propriétés semi-conductrices, photo-conductrices et photo-voltaïque.

Le sélénium est un élément rare, présent très souvent à l'état de traces dans les sulfures naturels où il se substitue au soufre. Les espèces minéralogiques qui en contiennent sont les sélénures de cuivre, de plomb, d'argent, de thallium et de mercure, mais elles sont

trop peu abondantes pour en constituer des minerais d'extraction. Aussi, la production de sélénium repose sur sa présence en tant que sous-produit de l'affinage électrolytique du cuivre, obtenu à partir du sulfure de cuivre.<sup>[1.5]</sup>

Le sélénium fut une des premières substances photoélectriques découvertes : la conductivité de ce semi-conducteur augmente avec son irradiation lumineuse, jusqu'à lui procurer une activité photo-voltaïque. Cette propriété explique son usage actuel en xérographie (technique de reprographie), dans des détecteurs numériques aux rayons X et dans des cellules photoélectriques ou photoconductrices. D'autres usages de cet élément sont aujourd'hui plus répandus. Il est notamment employé en métallurgie pour améliorer la machinabilité, la résistance à la corrosion et la finesse de surface des aciers. Il peut également former des alliages avec la plupart des métaux.

Usages industriels	Proportions
Industrie du verre	50%
Chimie et fabrication de pigments	15%
Métallurgie	13%
Applications électriques	10%
Applications agricoles et biologiques	4%
Autres (vulcanisation...)	8%

**Tableau 2** – Usages industriels du sélénium et de ses composés.<sup>[1.6]</sup>

Le pigment rouge de cadmium est, quant à lui, du sulfoséléniure de cadmium, utilisé dans les encres rouges. L'industrie du verre, qui est la plus grande consommatrice de sélénium, l'utilise comme colorant, de jaune à rouge, ou comme décolorant, par compensation de la couleur verdâtre des impuretés ferriques. En agriculture, il est additionné aux engrais sur des sols pauvres en sélénium et dans l'alimentation animale pour augmenter la croissance du bétail et des volailles, et pour en prévenir certaines pathologies. Le sélénium a également des usages en chimie organique. La production moyenne de sélénium, pour répondre à ces besoins, est de l'ordre de 2000 tonnes par an.<sup>[1.6]</sup>

## 2 Toxicité et déficience

L'utilisation industrielle usuelle de cette substance ne doit pas en faire oublier le danger et la toxicité. Celle-ci peut, en effet, être à l'origine de pollutions sévères. Bien sûr, chez les animaux, la toxicité aiguë du sélénium appelée "blind staggers", et son rôle dans certaines maladies comme l'épizootie sèche ou "alkali disease", sont connus depuis les années 30.<sup>[1.7]</sup> Mais ce n'est qu'à partir des années 80 qu'a été réalisée l'importance de l'enjeu environnemental que constitue le contrôle de la propagation de ce polluant, lorsque le réservoir naturel de Kesterson, Californie, devint fortement pollué par des sélénates provenant des eaux d'écoulement et d'irrigation agricole intensive. Cette contamination

(plus de 1400  $\mu\text{g}$  de sélénium total par litre d'eau mesuré en 1982) provoqua une forte bio-accumulation de sélénium et donc une mortalité importante chez les oiseaux migrateurs, poissons, insectes, algues et plantes peuplant ce réservoir naturel.<sup>[1.8]</sup>

### Non-discrimination entre le soufre et le sélénium

Le taux de sélénium dans le vivant, animal ou végétal, est parallèle à son abondance dans le sol ou la nourriture. Ceci découle du fait que la plupart des étapes de l'assimilation du soufre, enzymatiques ou chimiques, ne sont pas discriminantes entre les deux composés aux propriétés très proches<sup>[1.9]</sup> : l'intégration du sélénium dépend entièrement de sa proportion dans les ressources alimentaires. Les organismes ne disposent donc pas de parade à la forte présence de cet élément dans leur milieu. Chez les animaux d'élevage, les conséquences d'une intoxication chronique au sélénium sont la perte de poids et de poils, la déformation des sabots, une raideur des articulations et une forte baisse de la vitalité<sup>[1.7]</sup> : le sélénium s'accumule dans les reins, le foie, les muscles et les extensions kératinisées.

Composés	Espèces	DL50 orale
Sélénium élémentaire	rat/cobaye	6700 mg/kg
Sélénite de sodium	rat/souris	7 mg/kg
	lapin	2,25 mg/kg
	cobaye	5,06 mg/kg
Séléniate de sodium	rat	1,6 mg/kg
	lapin	2,25 mg/kg

**Tableau 3** – Données de toxicité aiguë des composés du sélénium. (INRS fiche n°150)

Une fois assimilé, la forte toxicité du sélénium est due également à son homologie avec le soufre. En effet, pratiquement tous les composés séléniés que l'on retrouve dans les plantes, bactéries, levures ou animaux sont des isologues de composés soufrés.<sup>[1.9]</sup> Les remplacements non-spécifiques du soufre par le sélénium dans les bio-polymères comme les protéines, les acides nucléiques et les dérivés complexes de glucides peuvent avoir des effets délétères pour les organismes dans lesquels ils surviennent.<sup>[1.10]</sup> Car bien que leurs propriétés soient proches, les composés séléniés sont moins stables et plus réactifs que les composés soufrés, ce qui semble compter fortement dans la toxicité du sélénium.<sup>[1.11]</sup> C'est le métabolisme du soufre qui est le plus touché par la forte présence de sélénium. Ainsi, parmi les enzymes qui utilisent sans les différencier le sélénium et le soufre, sont à noter l'ATP sulfurylase, l'ARN<sub>t</sub> sulfur-synthétase, la cystéinyl-ARN<sub>t</sub><sup>Cys</sup> transférase, la méthionyl-ARN<sub>t</sub><sup>Met</sup> transférase, la S-adenosylméthionine synthase et la S-adenosylméthionine méthyl-transférase.<sup>[1.7]</sup>

Les voies métaboliques des processus de méthylation, celles relatives aux acides aminés soufrés et aux acides nucléiques soufrés et la synthèse protéique sont donc affectées par la présence, en forte quantité, de sélénium.

## AJR et disponibilité du Sélénium

L'image positive accordée aujourd'hui au sélénium est donc à relativiser. D'autant plus que le sélénium n'a été que tardivement reconnu comme oligo-élément essentiel au développement et à la vie : en 1954 pour certaines bactéries,<sup>[1.12]</sup> en 1957 pour les oiseaux<sup>[1.13]</sup> et les mammifères.<sup>[1.14]</sup>

L'ambivalence du sélénium doit d'autant plus être gardée à l'esprit que la fenêtre entre déficience et toxicité est assez faible. Par exemple, les recommandations d'apport minimum pour les animaux sont de 50 à 100  $\mu\text{g}$  de sélénium par kg de fourrage (matière sèche) alors que l'exposition à des niveaux de 2 à 5 mg par kg de fourrage entraîne l'apparition des signes d'intoxication chronique.<sup>[1.8]</sup> Pour l'homme, les recommandations d'apports journaliers de l'Agence Française de sécurité sanitaire des aliments (AFSSA) sont de 50 à 80  $\mu\text{g}$  par jour. Il est fortement déconseillé de dépasser 2 mg par jour. À l'inverse, un apport de moins de 30  $\mu\text{g}$  par jour est considéré comme insuffisant, entraînant une carence pouvant avoir comme symptômes une grande faiblesse, voire une cardiomyopathie sévère (maladie de Keshan).

Si l'on observe les variations des apports alimentaires moyens en fonction des régions du globe, on s'aperçoit que le caractère sélénoprive ou sélénifère d'un territoire peut avoir une influence déterminante sur le santé de sa population. Dans les zones sélénoprives (Chine, Nouvelle-Zélande...), où les apports moyens sont de 20-30  $\mu\text{g}$  par jour, des carences peuvent être constatées, tandis que dans les zones sélénifères (Venezuela, Mexique, Canada, Japon, Brésil, Colombie, Chine), la population bénéficie d'apports moyens de l'ordre de 100-200  $\mu\text{g}$  par jour. Pour la plupart des pays européens, les apports nutritionnels moyens sont de l'ordre de 50-100  $\mu\text{g}$  par jour, avec 60  $\mu\text{g}$  par jour en France.<sup>[1.5]</sup> L'augmentation de la circulation des denrées alimentaires entre régions tend aujourd'hui à tempérer ce déterminisme géographique.

La plus grande partie de l'apport de sélénium alimentaire provient des plantes, en particulier des céréales. Ceci laisse à réfléchir puisque contrairement aux mammifères, bactéries et archaebactéries qui ont besoin de cet oligo-élément, la nécessité de sélénium pour la croissance des végétaux supérieurs reste controversée.<sup>[1.8]</sup>

## 3 Les végétaux et le sélénium

Les besoins des végétaux en oligo-éléments pour croître sont à mettre en relation avec leur utilisation en tant que cofacteurs dans l'activité enzymatique de l'organisme. Jusqu'à récemment, aucune preuve n'existait pour soutenir l'hypothèse de l'utilisation de sélénium dans des enzymes d'organismes du "règne vert". Or des protéines purifiées montrant une activité glutathion-péroxydase sélénium-dépendante ont été décrites pour *Chlamydomonas reinhardtii*,<sup>[1.15-1.17]</sup> pour les germes d'orge<sup>[1.18]</sup> et pour *Aloe vera*.<sup>[1.19]</sup>

De même, certains ARN<sub>t</sub> végétaux contiendraient également spécifiquement du sélénium, comme dans les germes d'orge.<sup>[1.20]</sup> Nous parlerons en détails des ARN<sub>t</sub> sélénisés à la page 30

Mais ces découvertes isolées ne permettent pas de conclure sur l'importance que revêt le sélénium pour les plantes. Bien sûr, toute plante contient du sélénium, puisque,

comme indiqué précédemment, les étapes de l'assimilation du soufre ne sont pas discriminantes entre les deux composés. Pour la plupart des végétaux, le taux de sélénium qu'ils contiennent est donc parallèle au taux de sélénium présent dans le sol. Les plantes ne sont néanmoins pas à égalité devant les phénomènes d'assimilation du sélénium : certaines sont des hyper-accumulatrices de cet élément. Les premières découvertes, appartenant à l'espèce *Astragalus*, ont été identifiées en 1935 au USA avec une concentration en sélénium 100 fois plus importante que dans les autres espèces végétales de la même zone. Cette particularité est supposée procurer à ces plantes un avantage de défense contre les herbivores, par leur toxicité.<sup>[I.21]</sup>

Les végétaux supérieurs peuvent donc à ce titre être divisés en trois catégories.<sup>[I.8]</sup>

- La plupart des céréales, herbes et fourrages sont des non-accumulateurs, et contiennent moins d'un maximum de 25 mg de sélénium par kg de matière sèche, selon la zone où ils sont implantés.
- Les hyper-accumulatrices accumulent jusqu'à plusieurs milliers de milligrammes de sélénium par kg de matière sèche, lorsqu'elles sont implantées dans des régions sélénifères. Dans cette catégorie peuvent être classés certains spécimens des espèces *Astragalus*, *Stanleya*, *Morinda*, *Neptunia*, *Oonopsis* et *Xylorhiza*, qui sont résistants aux fortes concentrations de sélénium dans le sol, et l'utilisent comme avantage défensif.
- Les accumulateurs secondaires, bien qu'implantés sur des sols à teneurs moyennes voire faibles en sélénium, sont capables de l'accumuler jusqu'à des concentrations de 1000 mg par kg de matière sèche. Des exemples de telles plantes sont certains représentants des espèces *Aster*, *Astragalus*, *Atriplex*, *Brassica*, *Castilleja*, *Comandra*, *Grayia*, *Grindelia*, *Gutierrezia*, et *Machaeranthera*.

Étant donné l'effet toxique du sélénium sur les plantes en général, on peut se demander comment ces plantes hyper-accumulatrices sont capables de tolérer une telle concentration de sélénium dans leurs cellules. Le principal mécanisme semble être la multiplication des espèces organiques séléniées présentes.<sup>[I.22]</sup> Cela permet de garder à un niveau assez bas les concentrations de séléno-méthionine et de séléno-cystéine, qui, sinon, seraient incorporés dans les protéines à la place de méthionines et de cystéines avec l'effet toxique que l'on peut prévoir sur les fonctions métaboliques vitales de la plante. Des acides aminés ou des dipeptides non-protéiques peuvent être accumulés comme, par exemple, de la Se-méthylséléno-cystéine, de la séléno-cystathionine ou de la  $\gamma$ -glutamyl-Se-méthylséléno-cystéine.<sup>[I.8]</sup> Certaines plantes isolent même ces composés en les restreignant à des vacuoles cytoplasmiques. Un mécanisme alternatif est l'exclusion du sélénium des protéines par la discrimination des séléno-aminoacides, comme chez *A. bisulcatus* où la cystéinyl-ARN<sub>t</sub> synthétase est une des seules capables de distinguer la séléno-cystéine de la cystéine, et de n'estérifier que cette dernière sur l'ARN<sub>t</sub><sup>Cys</sup>.<sup>[I.23]</sup>

Bien sûr, ces plantes accumulatrices peuvent jouer un rôle important sur deux plans : la bio-remédiation<sup>[I.24]</sup> des problèmes de pollution au sélénium dans les zones contaminées, et la disponibilité d'une source de sélénium transportable pour les régions sélénioprives, dans le but de combler le déficit de cet élément dans leur sol.

## 4 Le sélénium dans nos cellules

Nous l'avons vu, la toxicité du sélénium est connue depuis longtemps, contrairement à la nécessité de la présence de cet élément dans notre alimentation. La découverte de ce besoin en sélénium vient de l'observation de certaines maladies touchant les animaux d'élevages.

En effet, certains syndromes pathologiques ont été identifiés comme étant des syndromes de déficience, pouvant être prévenus et quelques fois traités par l'administration de petites doses de sélénium.<sup>[1.7]</sup> Parmi ces maladies, on peut dénombrer des dystrophies musculaires connues sous le nom de "white muscle disease" affectant les agneaux et les veaux, des œdèmes important pouvant devenir hémorragiques appelés "exudative diathesis" chez les volailles, des nécroses dégénératives du foie ou des fibroses pancréatiques.

Les actions bénéfiques qu'avait la supplémentation en sélénium sur ces maladies engendrèrent un nouvel intérêt pour cet élément et pour la compréhension de son action et la découverte de sa forme biologique. C'est ainsi qu'en 1973 fut découverte la fixation du sélénium par certaines protéines. Les premières sélénoprotéines à être alors identifiées furent une glycine réductase,<sup>[1.25]</sup> une formate déshydrogénase<sup>[1.26]</sup> et une glutathion peroxydase.<sup>[1.27,1.28]</sup> Cette dernière était déjà caractérisée comme étant une enzyme responsable de la protection, chez les mammifères, de la membrane des globules rouges et d'autres tissus fonctionnant dans un environnement hautement aérobie, contre les dommages dus aux peroxydes organiques et aux hydroperoxydes.<sup>[1.29,1.30]</sup> Ce mécanisme de défense cellulaire est essentiel pour la survie animal, et joue également divers rôles régulateurs dans le métabolisme énergétique de la cellule.<sup>[1.31,1.32]</sup> C'est donc depuis la découverte de son importance dans le fonctionnement de la glutathion peroxydase, que le sélénium a obtenu son image d'antioxydant protecteur des cellules.

### Le sélénium et le cancer

Bien avant la connaissance du rôle le sélénium dans la protection des cellules, des études épidémiologiques effectuées aux États-Unis suggéraient déjà un lien entre alimentation déficiente en sélénium et l'augmentation de la survenue de cancer, comme de la mortalité due à cette maladie.<sup>[1.33,1.34]</sup> Mais l'importance de la glutathion peroxydase dans la détoxification des hydroperoxydes, composés aux fortes propriétés mutagènes, augmenta la plausibilité du lien entre sélénium et cancer. De très nombreuses études expérimentales et cliniques révèlent néanmoins que le problème est bien plus complexe.<sup>[1.35-1.38]</sup>

Un bilan assez largement admis peut être résumé par les points suivants<sup>[1.9]</sup> :

- Des études animales sur des carcinogènes chimiques montrent qu'une faible activité de la glutathion peroxydase, liée à un apport alimentaire réduit en sélénium, augmente le risque d'initiation de cancers. Néanmoins, dans la plupart de ces études, un effet préventif significatif n'est observé qu'avec une supplémentation en sélénium bien au-delà de celui nécessaire à une activité optimale de la glutathion peroxydase.



- Un effet anticarcinogène similaire est observé chez les animaux sans déficience en sélénium, c.-à-d. avec une expression de sélénoprotéines optimum.
- La nature du composé sélénié apporté est d'une importance critique. Parmi les composés naturels, les composés Se-monométhylés se sont montrés particulièrement actifs.
- Des composés séléniés synthétiques, comme le triphénylsélénonium, qui ne sont pas des ressources en sélénium pour l'organisme, ont également une activité anticarcinogène.
- Les essais cliniques sur l'action préventive du sélénium, tel celui démontrant une réduction des carcinomes hépatiques dans une région sélénioprive chinoise,<sup>[1.39]</sup> démontrent simplement qu'un retour à la normale du taux de sélénium est suffisant.
- Une des seules études cliniques effectuées à grande échelle et en double-aveugle, montra l'absence d'incidence d'une supplémentation en sélénium de 200  $\mu\text{g}$  par jour sur l'occurrence de cancer de la peau chez une population sans séléno-déficience, mais mit en évidence une réduction de 63% du nombre de cas de cancer, tout type confondu, et de 50% de la mortalité liée, par cette supplémentation.<sup>[1.40]</sup>

L'ensemble de ces résultats s'accordent avec la possibilité qu'un apport insuffisant en sélénium engendre une augmentation des risques de mutagenèse, résultant en une initiation de la transformation de cellules en lignées malignes. Quant à l'action préventive du sélénium à forte dose sur les attaques de substances carcinogènes, de nombreuses hypothèses sont en débat. Son action pourrait découler : d'une altération du métabolisme des carcinogènes, d'un renforcement de la réponse immunitaire, de l'inhibition spécifique d'enzymes thiolées ou séléniées, d'une interférence avec la régulation du cycle et de la prolifération cellulaire, de l'inhibition de l'angiogenèse ou de l'induction de l'apoptose. Mais ces actions ne peuvent être reliées aux rôles des différentes sélénoprotéines connues, et pourraient être le résultat d'autres formes du sélénium.

## Les autres applications médicales

Parmi les applications sanitaires diverses du sélénium, citons rapidement celles qui ne découlent pas de l'intégration du sélénium dans les mécanismes enzymatiques.

Ainsi, un composé du sélénium, que nous n'avons pas encore évoqué, à des propriétés antimycosiques intéressantes. Il s'agit du sulfure de sélénium, de formule  $\text{SeS}_2$ , particulièrement actif sur *Malassezia furfur*, levure lipophile, agent du pityriasis versicolor. Cette mycose fréquente touche les zones séborrhéiques comme le cuir chevelu, le visage et le haut du thorax. Le sulfure de sélénium a également une action bénéfique sur la dermatite séborrhéique. Cet agent se trouve donc sous forme de crème et de shampoing.

À côté de cet emploi externe de sélénium, un nombre important d'études montrent une implication du sélénium dans l'amélioration de l'état de patients dans certaines pathologies, lors d'apports de doses modérées de sélénium. D'autres études montrent une action préventive de cet élément sur certaines pathologies, lors de complémentation du régime alimentaire en sélénium.

Voici listées les pathologies pour lesquelles des données sont disponibles :

- Dans le cas d'infection virale, comme dans celui du cancer, un niveau élevé de sélénium améliorerait la réponse du système immunitaire.<sup>[I.41]</sup> À l'inverse, un taux de sélénium sanguin faible altère la réponse immunitaire de souris contre le virus influenza.<sup>[I.42]</sup> Une complémentation provoque une meilleure réponse immunitaire.
- Pour des virus produisant un stress oxydant important pour l'organisme, comme les virus des hépatites A, B et C, cette observation est d'autant plus pertinente. Lors de l'infection par l'un de ces derniers, le niveau des défenses de l'organisme contre le stress oxydatif diminue, comme la concentration sanguine en sélénium. Il y aurait donc un intérêt à maintenir le taux de sélénium et d'autres antioxydants à des niveaux supérieurs à la normale, en parallèle aux traitements classiques de ces infections.<sup>[I.43]</sup>
- Dans le cas d'une infection par le VIH, comme pour les virus hépatiques, l'organisme subit un stress oxydant important : d'où également la possibilité d'un appui au traitement classique du SIDA par un apport important d'antioxydant, dont le sélénium. Par contre, une étude clinique originale<sup>[I.44]</sup> va plus loin en prétendant que l'apport de sélénium améliore significativement l'état des patients. Cette étude, effectuée en double aveugle, rapporte des résultats controversés<sup>[I.45-I.47]</sup> : la supplémentation en sélénium ralentirait la progression du VIH et ferait remonter le taux de lymphocytes CD4. Il a en outre été montré que l'infection par le virus provoquait une chute du taux de sélénoprotéines dans les lymphocytes T et une augmentation de la concentration de composés séléniés de faible masse moléculaire.<sup>[I.48]</sup>
- Les maladies auto-immunes ont également fait l'objet d'études sur l'effet d'une complémentation en sélénium. Le cas de la thyroïdite auto-immune est particulièrement documenté, sans doute parce que le sélénium est connu pour être l'élément actif de certaines enzymes thyroïdiennes, les iodothyronine-déiodinases. La production d'hormone thyroïdienne active est donc dépendante de la présence de sélénium. De plus, les maladies auto-immunes sont généralement dues à un dérèglement du système immunitaire, dans lequel le sélénium a un rôle clé. Certaines études montrent que dans le cas de la maladie de Graves-Basedow, les apports de sélénium améliorent la compétence de la réponse immunitaire.<sup>[I.49]</sup> La baisse de la concentration des anticorps dirigés contre la peroxydase thyroïdienne (TPO-Ab) avait déjà été observée lors de co-traitement au sélénium de thyroïdite auto-immune.<sup>[I.50]</sup> La thyroïdite auto-immune chronique de Hashimoto a également été l'objet d'une récente étude clinique obtenant des résultats similaires.<sup>[I.51]</sup> Mais ces résultats sont contredits par une autre étude montrant que, dans le cas général, le niveau des marqueurs immunologiques des patients atteints de thyroïdite auto-immune n'est pas affecté significativement par un apport de sélénium.<sup>[I.52]</sup>
- Les patients en soins intensifs, sujets à une septicémie et à un syndrome sévère d'inflammation systémique, ont également bénéficié de supplémentation en sélénium. Les études cliniques menées sur ces cas montrent la remontée des taux du sélénium plasmatique et de l'activité glutathion-péroxydase, entraînant une réduction significative des défaillances généralisées des organes et même une réduction de la mortalité des patients les plus sévèrement touchés.<sup>[I.53]</sup> Une étude montre même que l'apport de fortes doses de sélénium

(1mg par jour) était un co-traitement bénéfique pour ces patients, avec une forte baisse de mortalité et un retour à la normale du taux sanguin de sélénium, sans effet secondaire.<sup>[I.54]</sup>

- En ce qui concerne la fertilité, le sélénium est un élément essentiel pour la spermatogénèse.<sup>[I.55]</sup> Une de ses utilisations prend place dans l'abondante synthèse d'une glutathion peroxydase (GPx4) spécifique aux spermatozoïdes. La GPx4 affiche même deux fonctions : sa pleine activité enzymatique est remplacée par une fonction structurale dans la capsule mitochondriale des spermatozoïdes matures.<sup>[I.56]</sup> Il a d'ailleurs été montré qu'un faible taux de sélénium provoque l'infertilité,<sup>[I.55]</sup> tout comme une mutation des sélénoprotéines GPx4 et P (SelP).<sup>[I.57]</sup>
- Le fonctionnement musculaire est également dépendant de la disponibilité en sélénium. Le développement musculaire et sa maintenance seraient liés à l'activité de certaines sélénoprotéines, dont la sélénoprotéine N<sup>[I.58]</sup> (SelN). Ainsi, une dégradation du fonctionnement de cette enzyme, de cause génétique ou pathologique, aurait le même effet qu'une carence en sélénium : une dystrophie ou une dégénérescence musculaire.<sup>[I.59]</sup>

Malgré des recherches intensives sur la compréhension de l'effet du sélénium sur ces maladies, la forme sous laquelle cet élément est biologiquement actif n'est pas toujours connue. Des composés sanguins du sélénium de faibles poids moléculaires sont supposés en partie responsables de l'action antioxydante observée lors d'apports sélénisés. Mais il est maintenant communément admis que la plupart des activités du sélénium sont dues à des protéines possédant un ou plusieurs atomes de sélénium.

Voyons plus en détail quels sont les différents membres de cette famille des sélénoprotéines et quelles sont les informations disponibles sur ces agents sélénisés de l'organisme.

## Les Sélénoprotéines

Ce sont donc des protéines synthétisées par l'organisme, contenant du sélénium. Cette liaison entre protéines et sélénium est connue depuis 1973.<sup>[I.25-I.28]</sup> Depuis, le nombre de sélénoprotéines connues et la quantité d'informations disponible sur ces protéines ont augmentés exponentiellement.

Sélénoprotéines	Abréviation
Formate déshydrogénase, sous-unité A	FdhA
Sélénophosphate synthétase	SelD
Hétérodisulfide réductase, sous-unité A	HdrA
Formylméthanofurane déshydrogénase, sous-unité B	FwdB
Hydrogénase F420-réductrice, sous-unité $\alpha$ ou $\delta$	FrhA ou FrhD
Hydrogénase F420-non-réductrice	vhuD et vhuU
Homologue à HesB	—

**Tableau 4** – Les sélénoprotéines connues chez les archaées.<sup>[I.60]</sup>

Elles se retrouvent dans les trois domaines du vivant, Eubacteria, Archaea et Eukaryota. Bien qu'il existe au moins une sélénoprotéine commune aux trois domaines : la sélénophosphate synthase (voir les tableaux 4, 5 et 6), la plupart sont spécifiques. De plus, il existe des organismes ne possédant pas de sélénoprotéines, également dans les trois domaines<sup>[1.61]</sup> : la branche des crenarchaeotes chez les archaées, diverses branches des eubactéries, comme les cyanobactéries, et chez les eucaryotes, les branches des champignons et des plantes, cette dernière information pouvant être révisée à l'éclairage de nouveaux éléments (voir § 3, page 21).

Sélénoprotéines	Abréviation	Activité
<i>Contenant une sélélocystéine</i>		
Formate déshydrogénase, sous-unité A	FdhA	métabolisme énergétique
Sélénophosphate synthétase	SelD	activation du sélénium
Peroxyrédoxine	Prx	régulation du statut redox
Thiorédoxine	Trx	”
Glutathion peroxidase	GPx	”
Glutarédoxine	Grx	”
Glutathion S-transférase	—	”
Dithiol:disulfure oxydoréductase	—	”
Glycine réductase, sélénoprotéine A	GrdA	métabolisme énergétique
Glycine réductase, sélénoprotéine B	GrdB	”
Proline réductase	PrdB	”
Hydrogénase F <sub>420</sub> -réductrice, sous-unité $\alpha$	FrhA	”
Hydrogénase F <sub>420</sub> -réductrice, sous-unité $\delta$	FrhD	”
NADH oxydase	—	”
Hétérodisulfure réductase, sous-unité A	HdrA	”
Fe-S oxydoréductase	GlpC	”
Arsénate réductase	—	”
Protéine de biosynthèse de molybdoptérine	MoeB	synthèse de cofacteur
S-transférase homologue à la rhodanèse	—	inconnue
Homologue à Prx	—	inconnue
Homologue à AhpD	—	inconnue
Homologue à SelW	—	inconnue
Homologues à DsbA, DsbE ou DsbG	—	inconnue
Homologue à HesB	—	inconnue
Homologue à OsmC	—	inconnue
<i>Contenant un sélénium labile</i>		
CO déshydrogénase	CoxL	métabolisme énergétique
Acide nicotinique hydroxylase	—	”
Xanthine déshydrogénase	—	”

**Tableau 5** – Les sélénoprotéines connues chez les eubactéries.<sup>[1.9,1.62]</sup>

Sélénoprotéines	Abréviation	Activité
<i>Glutathion peroxydases</i>	GPx	
GPx cytosolique ou classique	cGPx, GPx-1	métabolisme énergétique
GPx gastrointestinale	GI-GPx, GPx-2	et
GPx du plasma	pGPx, GPx-3	régulation du statut redox
Phospholipide-hydroperoxyde GPx	PHGPx, GPx-4	
<i>Iodothyronine déiodinases</i>		
5'-Déiodinase, type I	5'DI, DI1, D1	
5'-Déiodinase, type II	5'DII, DI2, D2	production et régulation
5-Déiodinase, type III	5DIII, DI3, D3	hormonale (thyroïdienne)
<i>Thiorédoxine réductases</i>	TrxR	
Thiorédoxine réductase 1	TR1	régulation du statut redox
TR testiculaire	TR2	et
TR mitochondriale	TR3	autres activités
TR homologues	SelZf1, SelZf2	
<i>Autres</i>		
Sélénophosphate synthétase-2	SPS2	activation du sélénium
Méthionine-S-sulfoxyde réductase	MsrA	réparation de protéines <sup>[I.63]</sup>
Sélénoprotéine BthD	BthD	inconnue (Drosophile) <sup>[I.64]</sup>
Sélénoprotéine G-riche	G-rich	inconnue (homologue à SelK <sup>[I.65]</sup> )
Sélénoprotéine H	SelH	oxydoréductase nucléaire <sup>[I.66]</sup>
Sélénoprotéine I	SelI	inconnue <sup>[I.67]</sup>
Sélénoprotéine J	SelJ	structurale (cristallin de poissons) <sup>[I.68]</sup>
Sélénoprotéine K	SelK	régulation du statut redox <sup>[I.69]</sup>
Sélénoprotéine L	SelL	homologue aux TrxR <sup>[I.70]</sup>
Sélénoprotéine M	SelM	modulation <sup>[I.71]</sup>
Sélénoprotéine N	SelN	développement musculaire
Sélénoprotéine O	SelO	inconnue <sup>[I.67]</sup>
Sélénoprotéine P	SelP, SepP	transport et régulation
Sélénoprotéine R (ou MsrB <sup>[I.63]</sup> )	SelR/SelX	méthionine-R-sulfoxyde réductase <sup>[I.63]</sup>
Sélénoprotéine S	SelS	inconnue <sup>[I.67]</sup>
Sélénoprotéine T	SelT	modulation <sup>[I.72]</sup>
Sélénoprotéine U	SelU	inconnue <sup>[I.73]</sup>
Sélénoprotéine V	SelV	homologue aux TrxR <sup>[I.74]</sup>
Sélénoprotéine W	SelW	homologue aux TrxR <sup>[I.74]</sup>
Sélénoprotéine de 15 kDa	Sel15/Fep15	inconnue <sup>[I.48]</sup>

**Tableau 6** – Les sélénoprotéines connues chez les eucaryotes.<sup>[I.75]</sup>

Un des mystères des sélénoprotéines, résidant dans le mode de liaison du sélénium aux protéines, a été très rapidement levé par la découverte de la présence d'une sélénocystéine dans la protéine A du complexe de la glycine réductase de *Clostridium stickandii* en 1976.<sup>[I.76]</sup> Cet acide aminé particulier sera vu en détail au chapitre 1, page 33.

Néanmoins, toutes les "séléno-protéines" n'ont pas de sélénocystéine : un petit groupe de protéines de bactéries possède un atome de sélénium labile, dont la fixation doit être différente (voir tableau 5, page 27). Les traitements de ces enzymes par des hautes températures ou des agents alkylants mènent à la libération de sélénium, sous forme de di-alkyle sélényde dans le dernier cas.<sup>[I.9]</sup> Ce sélénium semble être lié à une molybdoptérine, cofacteur des enzymes concernées.<sup>[I.10]</sup>

Loin de ces exceptions, la quasi-unanimité des sélénoprotéines possède donc un sélénium actif sous forme de sélénocystéine. Le dénombrement de ces sélénoprotéines a été l'objet d'une course à la découverte, avec, entre autres, les équipes de Vadim N. Gladyshev, Dolph L. Hatfield et Marla J. Berry travaillant sur le sujet. Les techniques modernes d'analyses informatiques des génomes séquencés ont permis de mettre à jour, dans les génomes connus, 7 familles de sélénoprotéines chez les archaées, 25 familles chez les eubactéries et plus de 30 familles chez les eucaryotes. Ces travaux ont pu, conjointement, faire naître une base de donnée nommée «SelenoDB 1.0»,<sup>[I.67]</sup> qui regroupe les données sur les sélénoprotéines connues chez les eucaryotes, avec comme prévision l'ajout futur de celles des procaryotes.

L'ensemble des sélénoprotéines présentes chez un organisme, ou sélénoprotéome, dépend grandement du type d'organisme considéré. Ainsi, il a été montré que, chez les eucaryotes, les grands sélénoprotéomes sont associés à des organismes aquatiques, alors que les organismes terrestres possèdent des sélénoprotéomes plus réduits.<sup>[I.77]</sup> La dynamique d'évolution des sélénoprotéomes semble donc très liée au niveau de ressource en sélénium disponible dans le milieu, celui-ci pouvant être considéré comme stable en milieu aquatique. Ces considérations sur l'évolution de l'usage de la sélénocystéine dans les protéines seront approfondies au § 1.6, page 67.

Parmi toutes les sélénoprotéines découvertes, un grand nombre d'entre elles ont une fonction encore inconnue à ce jour. Dans celles qui ont été étudiées plus en détail, on retrouve bien sûr les très connues glutathion peroxydases, thiorédoxine réductases et sélénophosphate synthases (plus de détails sur cette dernière § 2.2.3, page 89), à cause de leur rôle primordial ou de leur découverte précoce. Plus récemment étudiées, les sélénoprotéines originales SelP et SelN ont montré un important intérêt médical. La première parce qu'elle fait partie des seules sélénoprotéines libres présentes dans le plasma, avec la GPx-3,<sup>[I.78]</sup> ce qui en fait un accessible biomarqueur du statut en sélénium des patients.<sup>[I.79]</sup> Cette sélénoprotéine glycosylée<sup>[I.80]</sup> est la protéine connue contenant le plus de sélénocystéine (jusqu'à 10 résidus chez l'homme) et ayant des liens directs avec l'approvisionnement de certains organes en sélénium.<sup>[I.57, I.81–I.86]</sup> La protéine SelN, quant à elle, a fait l'objet de recherches démontrant son importance pour le développement et le maintien musculaire, et son lien avec certaines maladies de dystrophie musculaire, lors de sa mutation.<sup>[I.58, I.59, I.87, I.88]</sup>

Avant de voir plus en détail au chapitre suivant la forme active du sélénium dans ces protéines, la sélénocystéine, et son mécanisme d'action (§ 1.5.2, page 63), terminons l'énumération des formes biologiquement importantes que prend spécifiquement le sélénium par les acides nucléiques séléniés.

## Les ARN<sub>t</sub> séléniés

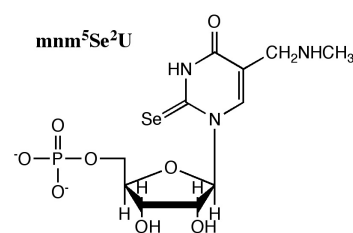
Il est intéressant de noter que les deux utilisations spécifiques du sélénium dans des bio-polymères ont comme point commun . . . le code génétique. En effet, comme nous le verrons en détail au chapitre qui lui est consacré (§ 1.2, page 35), le code génétique est une grille de lecture nécessaire pour traduire l'information génétique en protéine fonctionnelle. Ce code fait donc correspondre une unité de lecture, un triplet de l'ARN messager ou codon, à une unité constitutive des protéines, un acide aminé.

Les protéines séléniées contiennent l'acide aminé sélénocystéine, mais celui-ci ne fait pas partie des 20 acides aminés canoniques codés dans le code génétique : son intégration nécessite un recodage, dont nous allons voir le principe § 1.3.2, page 43.

Les ARN<sub>t</sub> sont, quant à eux, les garants du code génétique, grâce à la complémentarité de leurs anticodons avec les codons de l'ARN<sub>m</sub>, permettant la lecture de l'information. Dans un tel système, le nombre d'ARN<sub>t</sub> disponibles devrait être égal au nombre de combinaisons de triplets correspondant à un acide aminé, soit 61 chez la majorité des organismes. Pour réduire ce nombre excessif d'ARN<sub>t</sub> nécessaires et ainsi la complexité du système, les organismes utilisent l'astuce qui consiste à modifier chimiquement une base de l'anticodon de l'ARN<sub>t</sub>. Certaines bases modifiées s'apparient avec plus d'un type de base : un ARN<sub>t</sub> contenant une de ces bases sur son anticodon peut reconnaître ainsi de 2 à 4 triplets différents, réduisant ainsi le nombre d'ARN<sub>t</sub> utilisés. Le nombre réel d'ARN<sub>t</sub> existant chez la plupart des archaées et chez certains eucaryotes est de 46, le minimum observé étant de 28, pour certaines bactéries.<sup>[1.89]</sup>

Le sélénium entre dans la constitution d'un nucléotide d'ARN modifié : la 5-méthylaminométhyl-2-sélénouridine (mnm<sup>5</sup>Se<sup>2</sup>U). Ce nucléotide se retrouve dans des espèces d'ARN<sub>t</sub><sup>Lys</sup>, d'ARN<sub>t</sub><sup>Glu</sup> et d'ARN<sub>t</sub><sup>Gln</sup>, à la première position de leurs anticodons, appelée "wobble position". Cette position se retrouve en face du troisième nucléotide du triplet de l'ARN<sub>m</sub>.

Les triplets correspondants à la lysine dans le code génétique sont AAA et AAG, ceux correspondants au glutamate sont GAA et GAG, et enfin CAA et CAG pour la glutamine. Or l'uridine modifiée mnm<sup>5</sup>Se<sup>2</sup>U interagit de façon identique avec les nucléosides adénosine et guanosine.<sup>[1.90]</sup> Les ARN<sub>t</sub><sup>Lys</sup>, ARN<sub>t</sub><sup>Glu</sup> et ARN<sub>t</sub><sup>Gln</sup> qui la contiennent en "wobble position" sont donc suffisants pour décoder les deux triplets disponibles pour chacun de ces deux acides aminés. Cette stratégie d'économie d'ARN<sub>t</sub> est très utilisée chez les eubactéries,<sup>[1.89,1.91]</sup> mais se retrouve également parmi les archaées<sup>[1.92,1.93]</sup> et les eucaryotes.<sup>[1.94]</sup>



**Figure 1** – Formule de la 5-méthylaminométhyl-2-sélénouridine.

L'analogue soufré de ce nucléotide, la 5-méthylaminométhyl-2-thio-uridine ( $\text{mnm}^5\text{S}^2\text{U}$ ), est le précurseur de la  $\text{mnm}^5\text{Se}^2\text{U}$ . Ce précurseur est également présent à la "wobble position" d' $\text{ARN}_t^{\text{Lys}}$ , d' $\text{ARN}_t^{\text{Glu}}$  et d' $\text{ARN}_t^{\text{Gln}}$ . Ces  $\text{ARN}_t$  peuvent également être utilisés pour l'incorporation du glutamate, de la glutamine et de la lysine, mais contrairement aux  $\text{ARN}_t$  sélénés, ils s'apparient plus facilement avec l'adénosine qu'avec la guanosine positionnée en troisième position des codons correspondants. Cette préférence a pour conséquence un taux d'incorporation de glutamate et de lysine inférieur à celui atteint lorsque les  $\text{ARN}_t$  sélénés sont utilisés.<sup>[I.90]</sup>

La réaction de remplacement ATP-dépendant du soufre par le sélénium se déroule sur ces  $\text{ARN}_t$  thiolés, et nécessite un donneur de sélénium, identifié en 1992<sup>[I.95, I.96]</sup> comme étant du sélénophosphate  $\text{SePO}_3^{3-}$  synthétisé chez les eubactéries à partir de sélénium minéral et d'ATP par l'enzyme SelD. (Nous retrouverons SelD et le donneur de sélénium  $\text{SePO}_3^{3-}$  pour la synthèse de sélénoprotéines aux § 1.3.1, page 41, et § 2.2, page 87.)

L'enzyme qui catalyse ce remplacement chez les eubactéries a été découverte en 1994<sup>[I.97]</sup> puis caractérisée plus précisément en 2004<sup>[I.98]</sup> : cette  $\text{ARN}_t$  2-séléno-uridine synthase, produit du gène *ybbB*, est homologue au domaine protéique rhodanese. Les agents de ce remplacement pour les eucaryotes et les archées ne sont pas connus à ce jour.

D'autres espèces d' $\text{ARN}_t$  sélénés ont également été mises à jour et étudiées,<sup>[I.99]</sup> comme des dérivés de la 2-séléno-uridine<sup>[I.93]</sup> et la 4-séléno-uridine.<sup>[I.100]</sup> Cependant, la pertinence biologique de ces espèces n'est pas encore connue.

---

Un nombre important d'études sont en cours sur les différentes utilisations biologiques avérées et supposées du sélénium dans les organismes des trois domaines du vivant. Les découvertes sur lesquelles sont basées les connaissances actuelles sont très récentes, les plus anciennes ne datant que d'une vingtaine d'années. Dans ce contexte de connaissances mouvantes, il est important de surveiller les nouvelles voies d'exploration vers l'inconnu. Des exemples récents de recherches dans ce sens sont la découverte d'animaux originaux n'utilisant pas la sélénocystéine, et donc les sélénoprotéines,<sup>[I.101]</sup> ou d'organismes possédant une partie de la machinerie permettant la manipulation du sélénium, sans pour autant utiliser de sélénoprotéines ou d' $\text{ARN}_t$  sélénés.<sup>[I.102]</sup>

Le domaine de recherche le plus actif reste celui des sélénoprotéines, et donc de l'utilisation de la sélénocystéine, acide aminé original présenté dans le chapitre suivant.



**Note :**

La Partie I de ce manuscrit est divisée en trois chapitres, dont le premier chapitre introductif précédent : il s'agit d'une concaténation bibliographique sur le sélénium, la sélénocystéine, puis sur les acteurs de la biosynthèse de la sélénocystéine, qui nous intéressent particulièrement dans cette étude.

Sans être exhaustive, cette partie bibliographique est particulièrement dense, et cela à cause de la connexion du présent sujet à une thématique aux multiples embranchements.

Même si cela a pour conséquence de différer l'entrée dans le vif du sujet, il a été choisi de suivre un plan canonique, car celui-ci permet de disposer de tous les éléments nécessaires à la compréhension des liens logiques employés dans le travail de doctorat, qui sera, lui, exposé dans la Partie II.

La section 1.6, par exemple, évoque les questions sous-jacentes liées à l'acide aminé particulier qu'est la sélénocystéine, considérations qui, si elles ne sont pas essentielles dans la première approche mécanistique du problème posé, replace le sujet dans sa problématique globale.

De même, la section 1.4 apporte des éléments de comparaison essentiels, utilisés dans la suite du manuscrit, mais peut être laissée de côté dans une première lecture pour être abordée au moment où lesdits éléments seront cités.

Plusieurs lectures sont donc possibles.

# Chapitre 1

## La sélénocystéine

### Sommaire

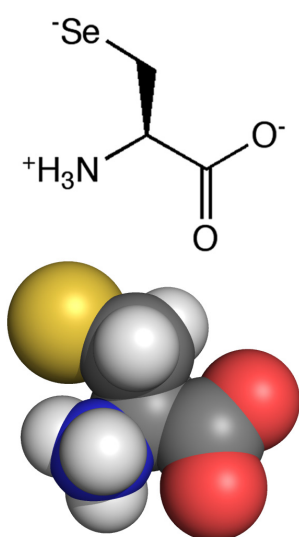
---

<b>1.1</b>	<b>Un acide aminé tardivement découvert . . . . .</b>	<b>34</b>
	présentation et historique . . . . .	34
<b>1.2</b>	<b>Rappels sur le code génétique et la traduction . . . . .</b>	<b>35</b>
1.2.1	La synthèse protéique . . . . .	35
1.2.2	Les 20 acides aminés canoniques : Le Code génétique . . .	38
<b>1.3</b>	<b>La sélénocystéine, exception du code génétique . . . . .</b>	<b>40</b>
1.3.1	Sa biosynthèse sur l'ARN <sub>t</sub> <sup>Sec</sup> . . . . .	41
1.3.2	Son incorporation co-translationnelle assistée . . . . .	43
	Cas des eubactéries, comparé aux eucaryotes et archaes .	43
1.3.3	Le cycle de l'acide aminé libre . . . . .	47
	ou le recyclage du sélénium par la cellule . . . . .	47
<b>1.4</b>	<b>Autres acides aminés particuliers comparables . . . . .</b>	<b>49</b>
1.4.1	L'asparagine et la glutamine . . . . .	49
1.4.2	La pyrrolysine, le 22 <sup>ème</sup> Acide Aminé . . . . .	51
1.4.3	La cystéine chez certaines Archaes . . . . .	55
<b>1.5</b>	<b>L'avantage catalytique de la sélénocystéine . . . . .</b>	<b>60</b>
1.5.1	Propriétés chimiques de la sélénocystéine . . . . .	60
1.5.2	Réactions enzymatiques impliquant une sélénocystéine . .	63
<b>1.6</b>	<b>Implications évolutives . . . . .</b>	<b>67</b>
1.6.1	Les sélénoprotéines Archaes . . . . .	68
1.6.2	Une grande disparité des sélénoprotéomes eucaryotes . . .	69
1.6.3	Hypothèses sur les facteurs entraînant cette disparité . . .	73
1.6.4	Confirmation de ces tendances par l'étude des eubactéries .	74

---

## 1.1 Un acide aminé tardivement découvert

### présentation et historique



**Figure 1.1** – Formule de la sélénocystéine.

La sélénocystéine est un acide aminé, similaire à la cystéine, mais dans lequel un atome de sélénium remplace l'atome de soufre. Cet acide aminé est symbolisé dans la nomenclature officielle par les trois lettres "Sec" et la lettre unique "U".<sup>[I.103]</sup>

Comme le sélénium est naturellement associé au soufre dans les minerais, avec un ratio<sup>[I.104]</sup> compris entre  $1/10^3$  et  $1/10^5$ , et du fait de leur proximité chimique, la sélénocystéine était longtemps considérée comme le produit de l'occurrence naturelle d'atome de sélénium dans les cystéines.

Car, la sélénocystéine était effectivement connue avant la découverte de son rôle dans les sélénoprotéines. Dès 1955,<sup>[I.105]</sup> des travaux montrent la présence de sélénium dans les protéines du plasma d'animaux de laboratoire, après injection de sels d'isotope radioactif  $^{75}\text{Se}$ . La forme sous laquelle ce sélénium est fixé à ces protéines est analysée en 1957<sup>[I.106]</sup> comme pouvant être des sélénocystéines et sélénométhionines, obtenues par l'introduction du sélénium dans le métabolisme naturel du soufre.

À l'état naturel, cet acide aminé a également été identifié sous la forme d'un métabolite méthylé, la Se-méthyl-sélénocystéine, par Shrift A. en 1963,<sup>[I.107]</sup> chez les plantes accumulatrices (cf § 3, page 21).

Lors de la découverte de la présence systématique de sélénium dans certaines protéines bactériennes et animales, en 1973 (cf § 4, page 23), de nombreuses hypothèses sur la fixation du sélénium ont été émises, dont celle de la présence du sélénium sous forme d'un composé séléno-organique.<sup>[I.108]</sup> Trois ans plus tard, l'entité séléniée contenue dans la protéine A du complexe de la glycine réductase de *Clostridium stickandii* est identifiée<sup>[I.76]</sup> : il s'agit d'une sélénocystéine.

Puis, en 1984, la détermination de la séquence d'acide aminé de la glutathion peroxydase (GPx) d'erythrocytes bovins permet de placer la sélénocystéine du site actif en 45<sup>ème</sup> position<sup>[I.109]</sup> de la chaîne protéique. À la surprise générale, le séquençage du gène de la glutathion peroxydase murine, en 1986,<sup>[I.110]</sup> montra la présence d'un codon UGA à la position de la sélénocystéine dans cette protéine, or il s'agit du codon terminateur opal, stoppant normalement la traduction.

À partir de cette découverte, les recherches sur ce phénomène biologique s'accroissent :

- En 1987, il fut montré que la sélénocystéine intégrée dans la GPx était dérivée du pool de sérine présent dans l'organisme, et non du pool de cystéine.<sup>[I.111]</sup>

- La même année, par l'étude de la traduction du gène de la formate déshydrogénase chez *Escherichia coli*, F. Zinoni montra que la traduction de la séquence en aval du codon UGA, présent à la 140<sup>ème</sup> position, était dépendant de la présence de sélénium dans l'organisme. Ce n'est plus le cas lorsque ce codon est muté en UCA (sérine) ou en UGC ou UGU (cystéine). On obtient alors une protéine entière, inactive dans le premier cas et, avec une activité moindre dans le deuxième cas. Indirectement, cela prouve que la sélénocystéine obéit, comme les acides aminés canoniques, à une incorporation cotraductionnelle opérée par le ribosome.
- Une telle découverte implique l'existence d'un ARN de transfert spécifique à la sélénocystéine, possédant un anti-codon UCA et portant l'acide aminé en question. Le gène d'un tel ARN<sub>t</sub><sup>Sec</sup> fut identifié en 1988,<sup>[1.3,1.4]</sup> en même temps que trois autres gènes codant pour les agents protéiques nécessaires à l'incorporation de l'acide aminé.

Depuis, la sélénocystéine est considérée comme le 21<sup>ème</sup> des acides aminés incorporés par le ribosome lors de la traduction. Pour comprendre en quoi cet acide aminé est exceptionnel, faisons un bref rappel sur la traduction protéique et le code génétique.

## 1.2 Rappels sur le code génétique et la traduction

### 1.2.1 La synthèse protéique

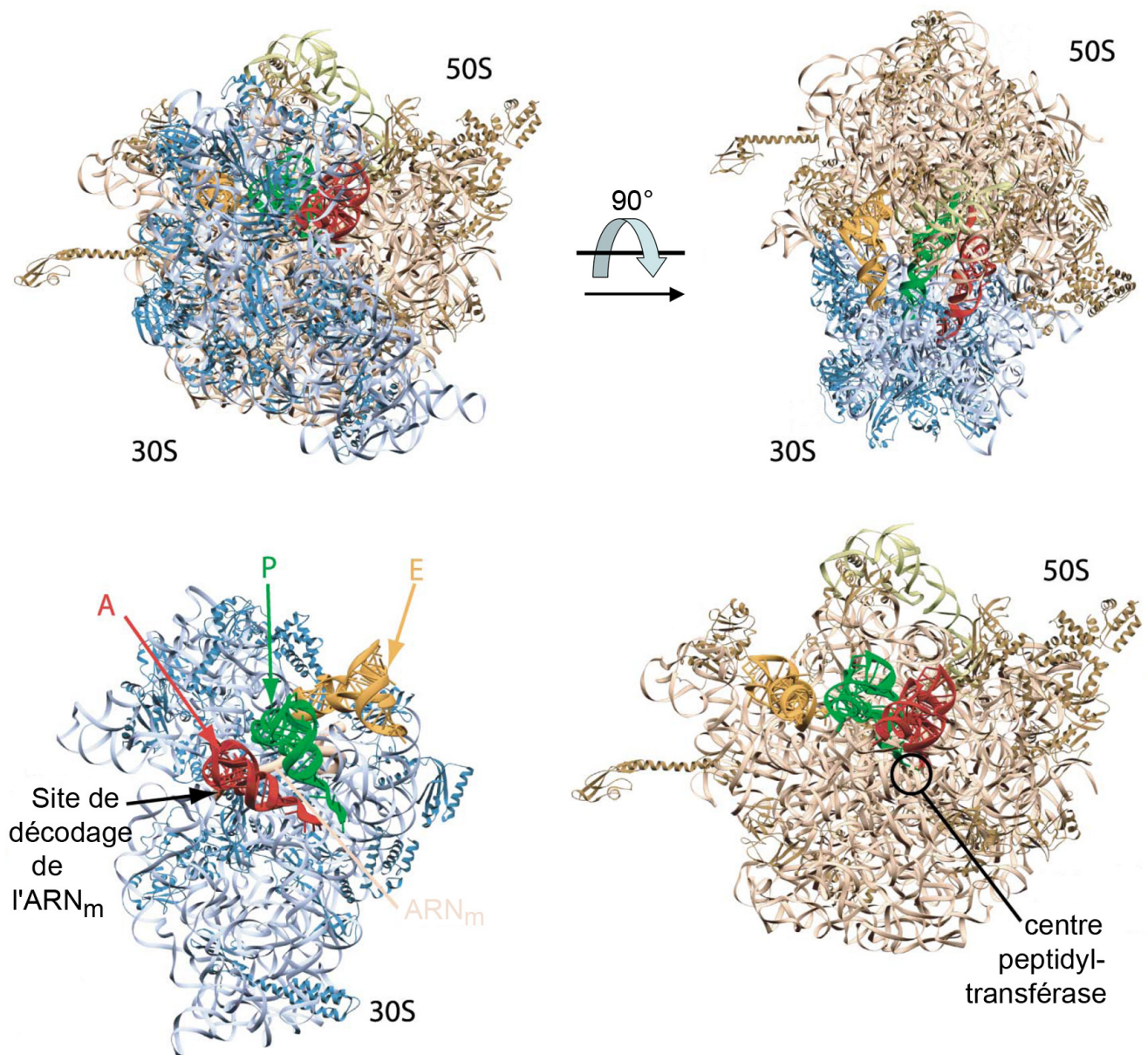
La synthèse protéique est guidée par l'information transportée par l'ARN<sub>m</sub>, provenant de la transcription de gènes inscrit dans la base de donnée qu'est l'ADN. Pour effectuer la lecture de l'ARN<sub>m</sub> et sa traduction exacte en séquence d'acides aminés, la cellule possède une complexe machine catalytique : le ribosome. Il est composé de plusieurs macromolécules d'ARN et de plus de cinquante protéines, interagissant pour former un complexe protéo-ribonucléique de plusieurs millions de Daltons.

Le ribosome 70S bactérien de 2,5 MDa se compose ainsi d'un tiers de protéines pour deux tiers d'ARN<sub>r</sub>, lui même divisé en deux sous-unités, 50S et 30S, nommées d'après leurs coefficients de sédimentation. La structure globale à haute résolution du ribosome bactérien a été obtenue en 2001 pour *T. thermophilus*<sup>[1.112]</sup> et en 2005 pour *E. coli*.<sup>[1.113]</sup> Une vue du 70S est présentée figure 1.2, page suivante.

	Ribosome bactérien	Ribosome eucaryote
Masse	2,5 MDa (70S)	4,2 MDa (80S)
Sous-unités	1,6 MDa (50S) & 0,9 MDa (30S)	2,8 MDa (60S) & 1,4 MDa (40S)
Contenu protéique	34 protéines & 21 protéines	49 protéines & 33 protéines
Taille (nucléotides)	~3000 nt & ~1500 nt	~5000 nt & ~1900 nt

**Tableau 1.1** – Données sur les ribosomes bactérien et eucaryote.

Bien que les deux ribosomes, bactérien et eucaryote, soient de tailles très différentes, ils sont très similaires dans leur agencement et leur fonction de traduction protéique. Voici la description des étapes du déroulement de la traduction, pour le ribosome bactérien :



**Figure 1.2** – Structure du ribosome bactérien avec ses protéines ribosomales, contenant trois  $ARN_t$  aux sites A, P et E.<sup>[1.114]</sup>

Lorsque le ribosome n'est pas en activité de traduction, ses deux sous-unités sont séparées. Lors de l'initiation de la traduction, catalysée par les trois facteurs IF1, IF2 et IF3, elles se rejoignent sur un  $ARN_m$  et forment le complexe de traduction en positionnant au niveau du site P le codon AUG initiateur de l' $ARN_m$  et l' $ARN_t^{fMet}$  complémentaire. Ce complexe ribosomal est alors prêt à effectuer le cycle d'élongation.

Ainsi le ribosome va sélectionner l'aminoacyl- $ARN_t$ , apporté par le facteur d'élongation EF-Tu, dont l'anticodon correspond au codon de l' $ARN_m$  présent dans son site A.

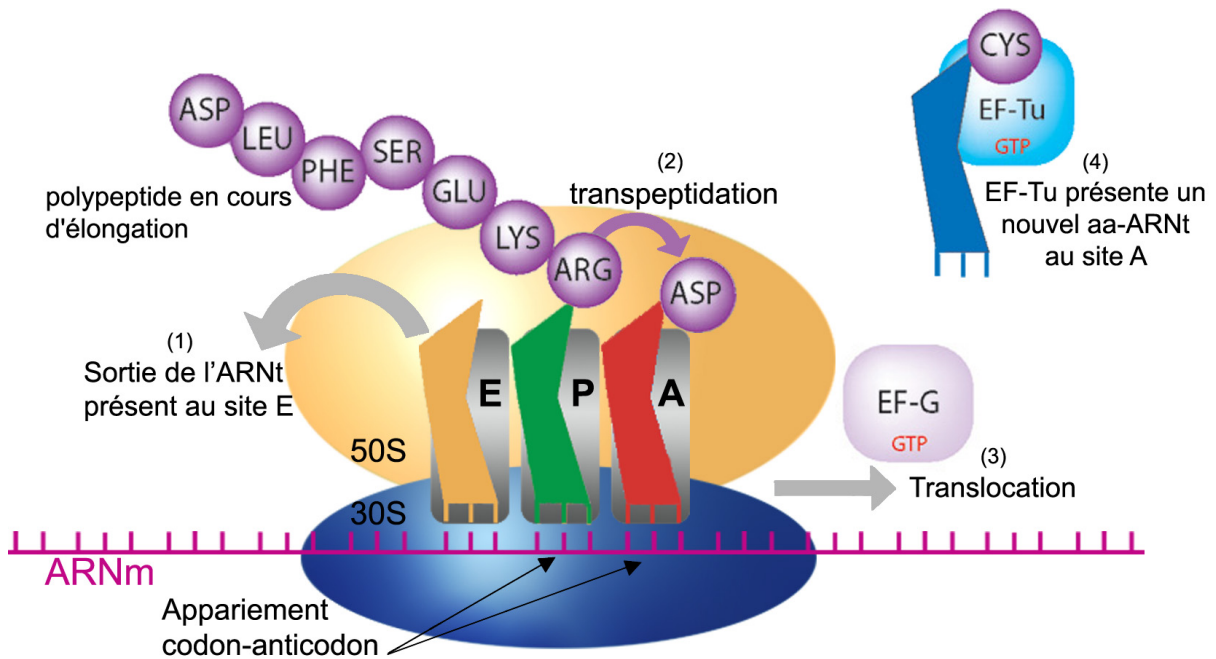


Figure 1.3 – Schéma du fonctionnement du ribosome bactérien.

L'acide aminé porté par son ARN<sub>t</sub> spécifique présent au site A et celui présent au site P se rejoignent au centre peptidyl-transférase de la sous-unité 50S et réagissent pour créer une liaison peptidique. Le facteur EF-G catalyse alors la translocation du ribosome d'un triplet sur l'ARN<sub>m</sub>, l'ARN<sub>t</sub> du site A se retrouvant déplacé au site P, coiffé de la chaîne protéique en formation, et celui du site P passant au site E. Le site A peut alors accueillir un nouvel ARN<sub>t</sub> apporté par EF-Tu et l'ARN<sub>t</sub> présent au site E sort du ribosome.

Le ribosome continue ainsi la traduction, de codon en codon, sur l'ARN<sub>m</sub>. Cette phase d'élongation de la protéine naissante, constituée de la répétition de ces étapes, est résumée par le schéma 1.3, ci-dessus.

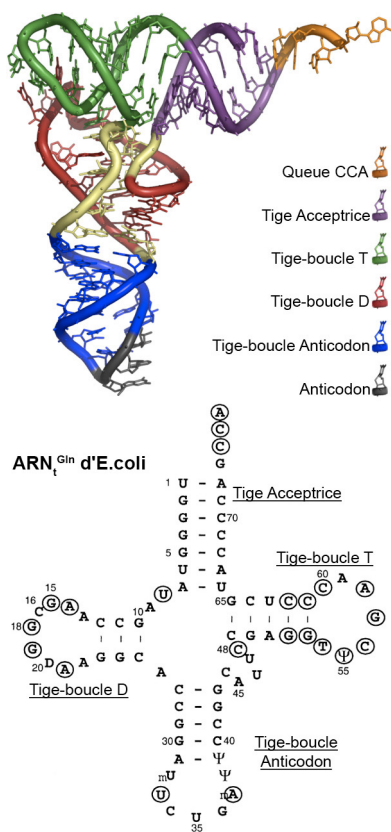
La synthèse protéique continue ainsi jusqu'à l'étape de terminaison engendré par la lecture d'un codon Stop (UAA, UAG ou UGA, chez presque tous les organismes). Un facteur de terminaison — RF1 ou RF2, selon le codon Stop rencontré — interagit alors au site A, provoquant l'hydrolyse et le relargage de la protéine. Le complexe est alors totalement dissocié par l'action successive des facteurs RF3 et RRF (facteur de recyclage du ribosome). Plus de détails sur les différentes étapes de la traduction sont disponibles dans les revues dont le sujet est l'analyse de ce phénomène.<sup>[I.114,I.115]</sup>

L'ajout d'acides aminés à la chaîne protéique en formation est effectué à la vitesse de 2 acides aminés par seconde pour le ribosome eucaryote, et jusqu'à 20 acides aminés par seconde pour le ribosome bactérien.<sup>[I.116]</sup> Malgré ces vitesses importantes, ils ont tout deux une excellente fidélité (environ 1 erreur tous les 10 000 acides aminés). L'efficacité et surtout la précision du ribosome, lors de cette "lecture" rapide du codon présent dans son site A, sont les garanties d'une traduction fidèle du message génétique en une protéine ayant la bonne séquence d'acide aminé.

Mais pour que cette lecture puisse avoir lieu, ce système de décodage spécifique nécessite un lien entre les deux polymères très dissemblables que sont la chaîne polypeptidique et le brin d'ARN<sub>m</sub>. Ce lien physique et logique est incarné par le pool des ARN de transfert spécifiques, dans lequel est inscrit l'abaque permettant un décodage sensé des gènes en protéines : le code génétique.

### 1.2.2 Les 20 acides aminés canoniques : Le Code génétique

Pour que les cellules synthétisent toujours des protéines identiques lors de la traduction d'un certain gène, un code fixe est nécessaire pour passer des triplets de l'ARN<sub>m</sub>, copie transformée de l'ADN, à la séquence des acides aminés constituant finalement la protéine. La nécessité d'un tel code fut comprise très tôt après la résolution de la structure de l'ADN, en 1953, par James D. Watson et Francis H. C. Crick. Ce dernier formulait déjà en 1955 l'hypothèse de l'existence d'un "adaptateur" entre les acides aminés de la protéine à construire et le polymère nucléique, celui-ci ne pouvant assurément pas être un patron direct et spécifique sur lequel les acides aminés libres interagissaient.



**Figure 1.4** – Exemple d'un ARN<sub>t</sub><sup>Gln</sup> : les bases cerclées sont invariantes ou semi-invariantes chez les ARN<sub>t</sub>.

Ces adaptateurs sont donc les ARN<sub>t</sub>, chaînes de 70 à 100 nucléotides en forme de L caractéristique, leur permettant de s'agencer dans les sites A et P du ribosome, découverts effectivement en 1958 par M. Hoagland et P. Zamecnik.<sup>[1.117]</sup> Leur structure tridimensionnelle spécifique est visible en figure 1.4, accompagnée de la représentation de leur structure secondaire en forme de feuille de trèfle, montrant les appariements au sein de la molécule.

Tous les ARN<sub>t</sub> possèdent des caractéristiques leur permettant d'être reconnu lors de la traduction :

- Une tige acceptrice, constituée de l'appariement sur 7 paires de bases des parties 5'-terminale et 3'-terminale de l'ARN<sub>t</sub>. Cette tige porte la queue non-appariée CCA en toute fin 3' de l'ARN, sur laquelle est estérifié l'acide aminé spécifique à l'ARN<sub>t</sub>.
- Une tige-boucle D et une tige-boucle T, comprenant respectivement jusqu'à 4 et 5 paires appariées.
- Une tige-boucle variable prenant place entre la tige-boucle anticodon et la tige-boucle T, inexistante pour les ARN<sub>t</sub> de Classe I comme celui donné en exemple, mais qui peut atteindre 20 nucléotides dans certains cas (comme pour des ARN<sub>t</sub><sup>Ser</sup>).
- Une tige boucle anticodon, portant ce triplet complémentaire au codon présent sur l'ARN<sub>m</sub> et comportant 5 paires de bases appariées. La "lecture" de l'ARN<sub>m</sub> se fait grâce à un appariement de Watson-Crick strict

entre les deux premières bases du codon de l'ARN<sub>m</sub> et les deux dernières de l'anticodon de l'ARN<sub>t</sub>, l'appariement de la position "wobble" respectant des règles plus lâches<sup>[I.115]</sup> pouvant faire intervenir un nucléotide modifié (voir § 4, page 30).

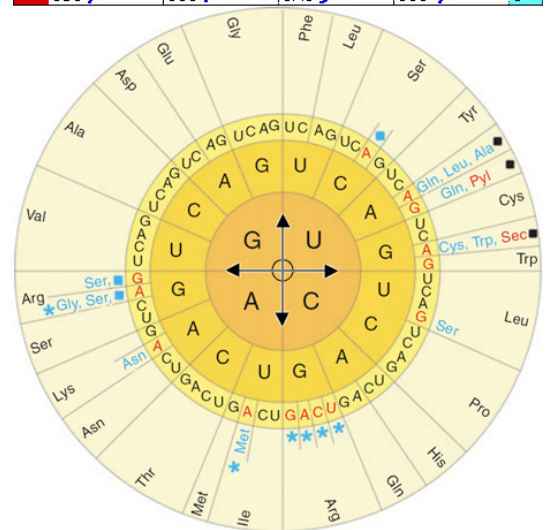
En portant un acide aminé spécifique et un anticodon propre, les ARN<sub>t</sub> représentent bien le code fixe, l'interface entre le message de l'ARN<sub>m</sub> et la constitution des protéines. Alors que l'anticodon est dépendant de la nature de l'ARN<sub>t</sub>, dont la séquence est inscrite dans un gène propre, l'estérification spécifique d'un acide aminé sur celui-ci dépend de l'action d'enzymes qui reconnaissent un pool d'ARN<sub>t</sub> et l'acide aminé libre qui leur est apparentés, pour les lier : les aminoacyl-ARN<sub>t</sub>-synthétases. C'est donc sur les couples spécifiques ARN<sub>t</sub> / aminoacyl-ARN<sub>t</sub>-synthétase que repose le code génétique.

Ce code, qui permet de prédire la séquence d'une protéine à partir de son ARN<sub>m</sub>, a été réellement déchiffré entre 1961 et 1966 par l'important travail fournit par Marshall W. Nirenberg et Har G. Khorana. Comme ce code paraissait semblable pour tout les organismes étudiés, il a été postulé que "l'accident" de sa création avait donné naissance à un code universel et gelé,<sup>[I.1]</sup> illustré par les schémas 1.5.

Mais l'augmentation exponentielle des connaissances disponibles sur un nombre grandissant d'organismes et sur les particularités de leur traduction, nous permet de savoir aujourd'hui que le dogme du Code universel n'est pas respecté : il possède de nombreuses exceptions et est encore en évolution.<sup>[I.118]</sup>

Selon les organismes, certains codons n'ont donc pas la même signification et d'autres ne sont pas utilisés. Cela est particulièrement le cas chez les organismes ou organites subissant de fortes pressions génétiques (comme les mitochondries). Ces réaffectations de codon illustrent un changement touchant l'ARN<sub>t</sub> de l'organisme en question : soit un changement de la nature du nucléotide présent en position "wobble", soit un changement de l'acide aminé estérifié à un certain ARN<sub>t</sub>. Un changement par modification chimique de cette première base de l'anticodon est courant, alors qu'un changement de fonctionnement de l'ainoacyl-ARN<sub>t</sub>-synthétase est plus rare.

1ère base	2nde base	U	C	A	G	3ème base		
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
	UUC		UCC		UAC		UGC	
	UUA		UCA		UAA		UGA	
	UUG		UCG		UAG		UGG	
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
	CUC		CCC		CAC		CGC	
	CUA		CCA		CAA		CGA	
	CUG		CCG		CAG		CGG	
A	AUU	Ile	ACU	Thr	AUU	Asn	AGU	Ser
	AUC		ACC		AAC		AGC	
	AUA		ACA		AAA		AGA	
	AUG		ACG		AAG		AGG	
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
	GUC		GCC		GAC		GGC	
	GUA		GCA		GAA		GGA	
	GUG		GAG		GAG		GGG	



**Figure 1.5** – Deux représentations du code génétique. La représentation circulaire contient les changements répertoriés chez certains organismes ou organites : Les carrés symbolisent des codons Stop, et les étoiles des codons pouvant être non-affectés.<sup>[I.118]</sup>



À ces exceptions, découlant d'un changement dans le code génétique de certains organismes, s'ajoute trois autres types de cas célèbres pour leurs particularités que nous allons comparer dans les sections suivantes :

- l'utilisation par certains organismes d'un nouvel  $\text{ARN}_t$  pour l'incorporation d'un nouvel acide aminé non-canonique, comme dans le cas de la sélénocystéine et de la pyrrolysine.
- l'utilisation d'une voie particulière pour obtenir la brique de synthèse protéique que constitue l' $\text{ARN}_t$  chargé de son acide aminé, comme dans le cas de de la sélénocystéine chez les organismes des trois domaines du vivant utilisant cet acide aminé, mais aussi de l'asparagine et la glutamine chez la plupart des eubactéries et archaes, et de la cystéine chez certaines archaes.
- Le cas décidément exceptionnel de la sélénocystéine, dont l' $\text{ARN}_t^{\text{Sec}}$  n'est pas reconnu par EF-Tu, et qui nécessite donc des facteurs protéiques propres et spécifiques pour pouvoir être incorporé lors de la synthèse protéique.

### 1.3 La sélénocystéine, exception du code génétique

Certains organismes possèdent donc, dans leur génome, le gène de l' $\text{ARN}_t$  supplémentaire permettant l'incorporation co-traductionnelle de la sélénocystéine. Comme précisé dans l'historique présent au début de ce chapitre, les différents gènes nécessaires à l'incorporation de la sélénocystéine chez les eubactéries ont été découverts en 1988,<sup>[1.3,1.4]</sup> dont en particulier le gène *selC* codant pour l' $\text{ARN}_t^{\text{Sec}}$ . Depuis, ses caractéristiques ont été déterminées pour un nombre important d'organismes.

La particularité la plus frappante de cet  $\text{ARN}_t^{\text{Sec}}$  est son anticodon UCA, complémentaire du codon Stop UGA. Cet  $\text{ARN}_t^{\text{Sec}}$  est appelé pour cela le supprimeur du codon opal ou "Opal-supprimeur". Néanmoins, chez les organismes utilisant la sélénocystéine et possédant l' $\text{ARN}_t^{\text{Sec}}$ , ce codon UGA conserve la possibilité d'avoir de façon ambiguë ces deux sens fondamentalement différents : soit l'arrêt de la traduction et le relargage de la protéine synthétisée, soit l'incorporation d'une sélénocystéine à la chaîne polypeptidique en formation et la poursuite de l'élongation de la protéine.

Cette ambiguïté est possible grâce à la nécessité, pour l'incorporation de la sélénocystéine, du recrutement message-dépendant d'un facteur d'incorporation spécifique remplaçant EF-Tu. Pour un codon UGA sans contexte particulier, les facteurs de terminaison de la traduction RF2 et RF3 officient normalement. Nous allons voir au § 1.3.2, page 43, la nature de ce contexte particulier et le fonctionnement de ce recodage.

Une difficulté supplémentaire s'oppose à un fonctionnement simplifié de l'incorporation de la sélénocystéine : le 21<sup>ème</sup> acide aminé ne possède pas d' $\text{ARN}_t$ -sythétase spécifique.

L'absence de sélénocystéinyl-ARN<sub>t</sub><sup>Sec</sup> synthétase peut s'expliquer par la nécessité vitale pour la cellule d'avoir une concentration interne en sélénocystéine libre la plus faible possible. En effet, la sélénocystéine et la cystéine étant difficilement discernées dans les voies métaboliques du soufre, la présence trop importante de sélénocystéine serait délétère pour la cellule. Cette toxicité, associée au problème de discrimination réciproque de la sélénocystéine et de la cystéine, n'a sans doute pas permis l'émergence d'une sélénocystéinyl-ARN<sub>t</sub><sup>Sec</sup> synthétase, chargée de recruter la sélénocystéine libre et de l'estérifier à son ARN<sub>t</sub><sup>Sec</sup>. La voie choisie par le vivant pour pouvoir utiliser tout de même cet acide aminé réactif sans danger est celle d'une biosynthèse sur son ARN<sub>t</sub><sup>Sec</sup>, exposée au paragraphe suivant.

La génération de protéines contenant des sélénocystéines implique que, lors de leur dégradation, de la sélénocystéine libre soit produite. Comme cela pose un problème de toxicité dans la cellule, il existe des voies efficaces de recyclage et de résorption de la sélénocystéine, que nous allons voir au § 1.3.3, page 47.

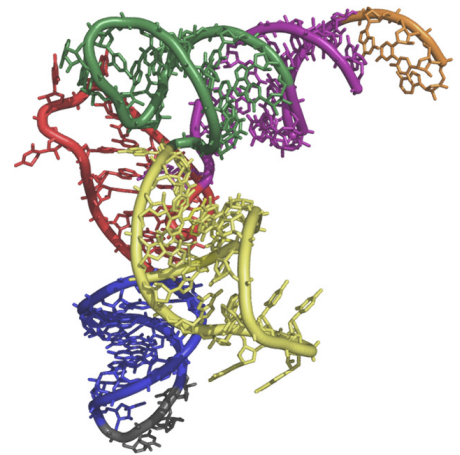
### 1.3.1 Sa biosynthèse sur l'ARN<sub>t</sub><sup>Sec</sup>

La biosynthèse de la sélénocystéine est précisément le thème d'étude de cette thèse. Nous nous intéresserons donc aux acteurs de la création du 21<sup>ème</sup> acide aminé, avec tout le détail nécessaire à l'acquisition des clés de la problématique abordée par ce travail, au chapitre 2. Mais voici tout d'abord un descriptif des différentes étapes nécessaires avant l'incorporation de la sélénocystéine dans les protéines.

Pour que la biosynthèse puisse avoir lieu sur l'ARN<sub>t</sub><sup>Sec</sup>, celui-ci doit passer par une étape préliminaire. La séryl-ARN<sub>t</sub>-synthétase (SerRS) reconnaît cet ARN<sub>t</sub><sup>Sec</sup> et estérifie une sérine à son extrémité 3', comme elle le fait sur le pool des ARN<sub>t</sub><sup>Ser</sup>, dédiés à l'incorporation de cet acide aminé. L'élément qui permet à la séryl-ARN<sub>t</sub>-synthétase de reconnaître tous ces ARN<sub>t</sub> est la présence d'une tige-boucle variable de longueur importante, élément commun des ARN<sub>t</sub><sup>Ser</sup> et ARN<sub>t</sub><sup>Sec</sup>. Un aperçu de la structure de la SerRS interagissant avec un ARN<sub>t</sub><sup>Ser</sup> est disponible figure 1.7, page suivante.

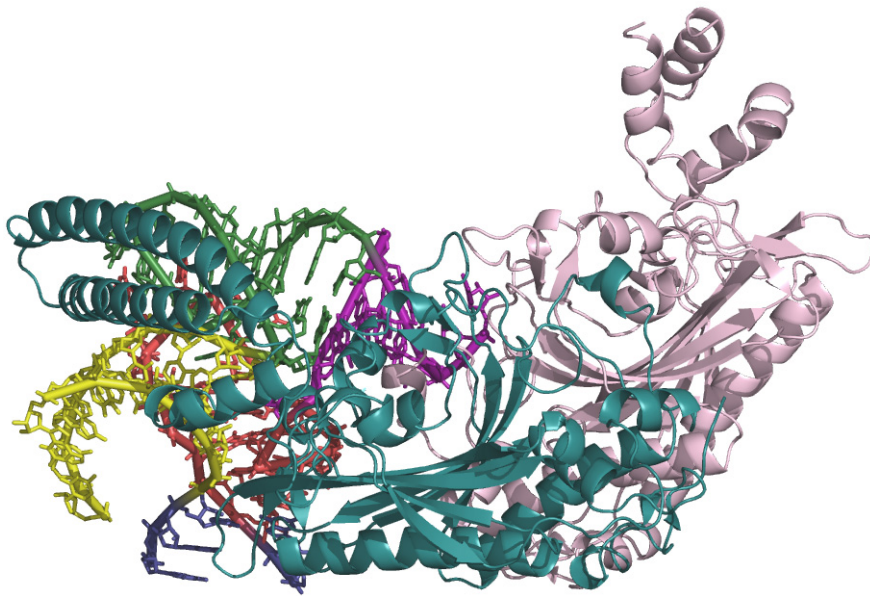
L'estérification de la sérine sur l'ARN<sub>t</sub><sup>Sec</sup> se fait néanmoins avec une efficacité moindre par rapport aux ARN<sub>t</sub><sup>Ser</sup>. De plus, alors que les séryl-ARN<sub>t</sub><sup>Ser</sup>, ou Ser-ARN<sub>t</sub><sup>Ser</sup>, sont reconnus par EF-Tu, le Ser-ARN<sub>t</sub><sup>Sec</sup> n'est pas reconnu, ce qui évite le risque d'une incorporation accidentelle d'une sérine aux codons UGA.

Le Ser-ARN<sub>t</sub><sup>Sec</sup> doit ensuite passer par une étape de transformation de la sérine estérifiée en sélénocystéine, étape effectuée différemment chez les eucaryotes

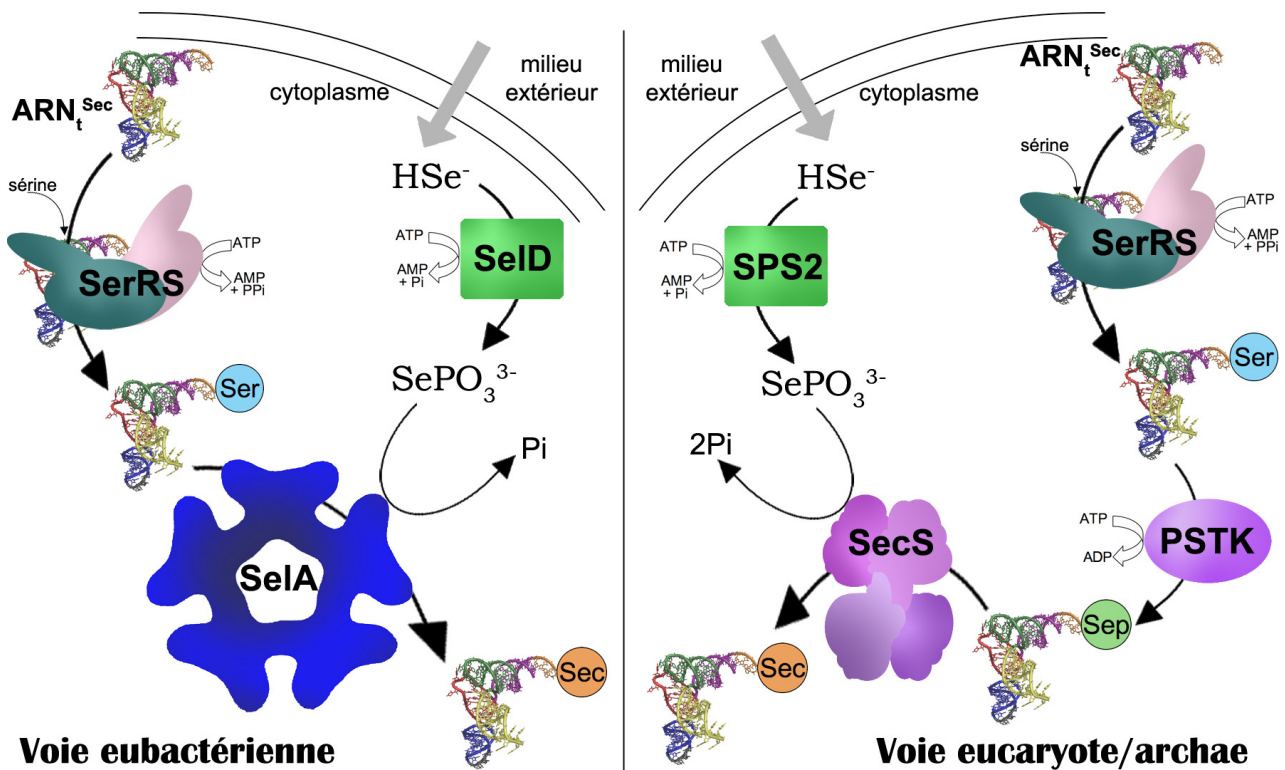


**Figure 1.6** – Modèle de structure tridimensionnelle de l'ARN<sub>t</sub><sup>Sec</sup>, obtenu grâce à des expériences d'empreintes chimiques et enzymatiques, et par homologie avec les structures connues.<sup>[I.119]</sup> Le code couleur respecte celui du schéma 1.4, page 38.

et chez les eubactéries.



**Figure 1.7** – Structure résolue du complexe entre un  $\text{ARN}_t^{\text{Ser}}$  et le dimère de la SerRS. [I.120] Les tiges acceptrice, anticodon et variable sont tronquées, car en partie absente de la densité électronique obtenue. Le dimère de SerRS interagit avec l' $\text{ARN}_t^{\text{Ser}}$  au niveau du site d'estérification et à la base de la tige-boucle variable. (N° PDB : 1SER)



**Figure 1.8** – Schéma représentant les voies eubactérienne et eucaryote/archae de biosynthèse du sélénocystéinyl- $\text{ARN}_t^{\text{Sec}}$ .

Ainsi, chez les eubactéries, cette biosynthèse du Sec-ARN<sub>t</sub><sup>Sec</sup> sur la base du Ser-ARN<sub>t</sub><sup>Sec</sup> se fait en une seule étape, catalysée par la sélénocystéine synthase bactérienne SelA.

Chez les eucaryotes et les archaes, le Ser-ARN<sub>t</sub><sup>Sec</sup> passe tout d'abord par une étape de phosphorylation, catalysée par la kinase PSTK,<sup>[1.121]</sup> puis le phosphoséryl-ARN<sub>t</sub><sup>Sec</sup> (ou Sep-ARN<sub>t</sub><sup>Sec</sup>) est transformée par une sélénocystéine synthase différente : la SepSecS. Le détail des connaissances disponibles sur ces sélénocystéine synthases et sur les mécanismes de réaction sera abordé au chapitre qui leur est consacré, page 90.

Dans les deux cas, l'atome de sélénium provient d'un donneur, le sélénophosphate SePO<sub>3</sub><sup>3-</sup>, synthétisé dans la cellule par la sélénophosphate synthétase à partir de sélénium minéral, issu du milieu extérieur. Cette enzyme est appelée SelD chez les eubactéries et SPS2 chez les eucaryotes et archaes. Chez ces derniers, il existe une autre enzyme sélénophosphate synthétase homologue d'identifiée, la SPS1, mais son rôle dans l'assimilation du sélénium a été remis en question.<sup>[1.101,1.122]</sup> De même, ces enzymes homologues seront décrites en détail à la section 2.2.3, page 89.

Le schéma 1.8 résume ces différentes étapes aboutissant au sélénocystéinyl-ARN<sub>t</sub><sup>Sec</sup>.

Ces deux voies de biosynthèse de la sélénocystéine sont très similaires, mais des informations structurales et enzymatiques sont en partie manquantes pour qu'une comparaison précise entre les différents acteurs de ces voies soit possible. L'intérêt d'une telle comparaison réside dans l'obtention de pistes supplémentaires quant à la filiation possible entre ces deux voies et une hypothétique voie ancestrale commune de synthèse de la sélénocystéine, préexistante, avant la divergence des trois domaines du vivant, correspondant à l'émergence de l'utilisation de cet acide aminé.

### 1.3.2 Son incorporation co-translationnelle assistée

Une fois le sélénocystéinyl-ARN<sub>t</sub><sup>Sec</sup> synthétisé, il peut être utilisé lors de la synthèse protéique. Mais comme indiqué précédemment, l'ARN<sub>t</sub><sup>Sec</sup>, quel que soit l'acide aminé esterifié à son extrémité 3', n'est pas reconnu par le facteur d'élongation classique de la synthèse protéique : EF-Tu (ou son homologue EF-1α chez les eucaryotes).

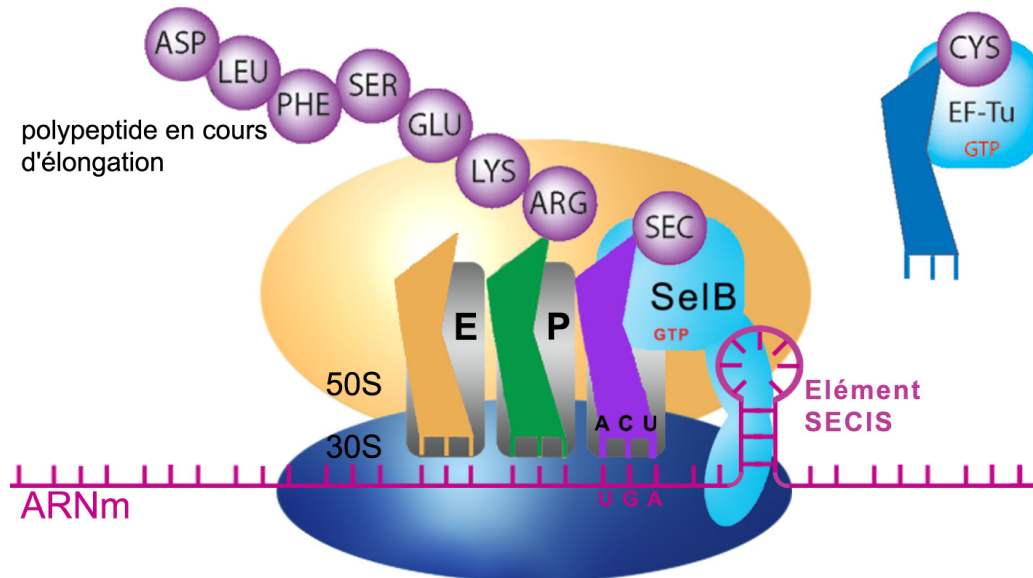
Cette particularité unique est nécessaire pour conserver l'ambiguïté du sens du codon UGA, pouvant être utilisé soit pour l'arrêt de la traduction, son utilisation majoritaire dans la plupart des génomes, soit pour l'incorporation d'une sélénocystéine et la poursuite de l'élongation de la protéine.

L'affectation d'un double sens à un codon nécessite l'utilisation d'un système de levée d'ambiguïté gène-dépendant : la machinerie traductionnelle doit adopter un comportement différent en fonction de l'ARN<sub>m</sub> traduit. Un message supplémentaire doit donc être disponible sur cet ARN<sub>m</sub>. Pour la sélénocystéine, il s'agit d'une structure secondaire de l'ARN, appelée SECIS pour "SElenoCysteine Insertion Sequence". Cette structure se situe directement en aval du codon UGA à recoder chez les eubactéries, alors qu'elle est présente dans la région 3' non-codant de l'ARN<sub>m</sub> chez les eucaryotes et les archaes.

Ce message est pris en compte par un facteur d'élongation spécifique, différent chez les eubactéries et chez les eucaryotes/archaes.

## La reconnaissance du signal et l'incorporation chez les eubacteries

Chez les eubactéries, un unique facteur protéique, SelB, interagit avec la structure secondaire SECIS de l'ARN<sub>m</sub> et avec le Sec-ARN<sub>t</sub><sup>Sec</sup>. Comme SECIS est situé immédiatement en aval du codon UGA à recoder, la disposition de l'ensemble permet au ribosome d'incorporer la sélénocystéine qui lui est présentée.<sup>[1.123]</sup>

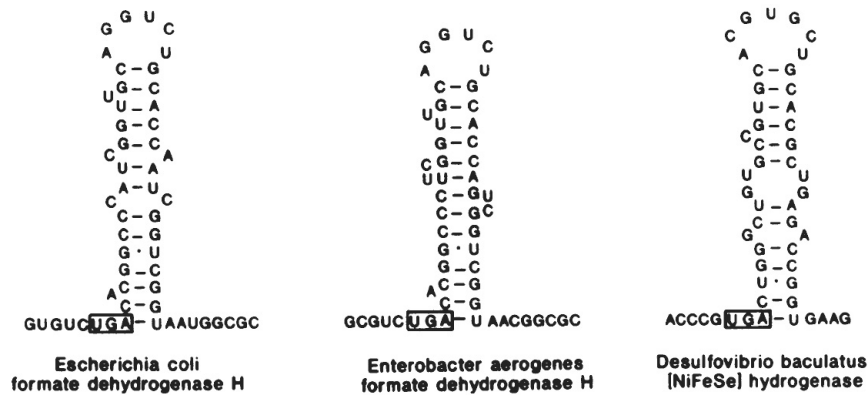


**Figure 1.9** – Schéma représentant, chez les eubactéries, la présentation au ribosome du Sec-ARN<sub>t</sub><sup>Sec</sup> par le facteur d'élongation spécifique SelB, reconnaissant la structure secondaire SECIS.

Les éléments SECIS bactériens sont composés d'une tige boucle d'une quarantaine de nucléotides. Chez *Escherichia coli*, les éléments essentiels à l'interaction avec SelB, et donc au recodage du codon UGA, sont les nucléotides G23 et U24 du sommet de la boucle, ainsi que l'uridine U17 en hernie, ou "bulge", que l'on retrouve le long de la tige.<sup>[1.124-1.126]</sup>

SelB est une protéine homologue à EF-Tu pour sa partie N-terminale, mais qui possède en C-terminal un domaine supplémentaire, composé de 4 motifs structuraux appelés "winged-helix", capables de lier l'ARN. L'affinité de SelB pour SECIS est grande, avec chez *Escherichia coli* un  $K_d$  de 30 nM pour la structure SECIS présente sur l'ARN<sub>m</sub> de la formate déshydrogénase H (FdhH). Cette affinité serait même améliorée d'un facteur 2 en présence de Sec-ARN<sub>t</sub><sup>Sec</sup>.<sup>[1.127,1.128]</sup>

Lorsque le ribosome arrive à proximité du codon UGA, la structure SECIS est dissociée sur ses 11 premiers nucléotides, la petite tige boucle résiduelle, de 17 nucléotides chez *E. coli*, étant la zone d'interaction avec SelB. L'incorporation de la sélénocystéine implique l'hydrolyse du GTP lié à SelB, ce qui provoque la libération de l'enzyme. La structure SECIS est par la suite entièrement dissociée pour permettre au ribosome de continuer la traduction jusqu'au codon Stop du gène traduit.



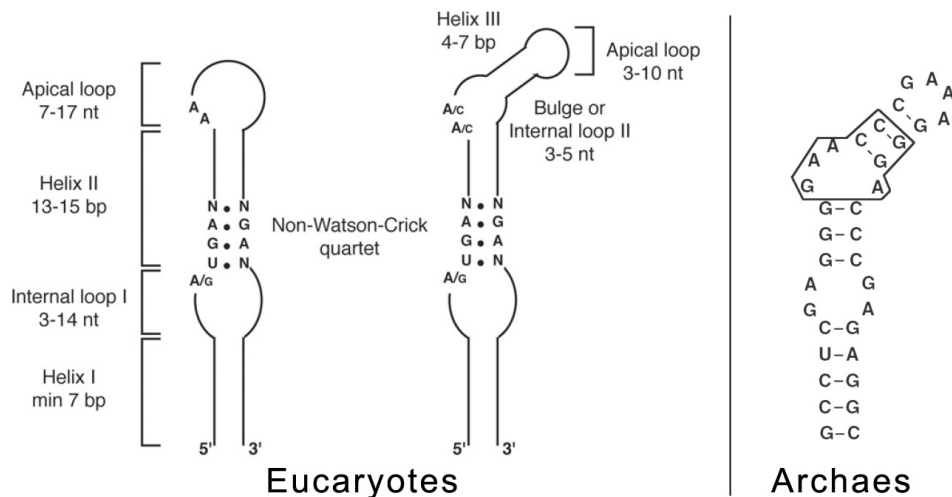
**Figure 1.10** – Structures SECIS retrouvées en aval du codon UGA codant pour la sélénocystéine, chez différentes bactéries.<sup>[I.124]</sup>

### Comparaison avec les eucaryotes et les archaés

Chez les eucaryotes et les archaés, le système de recodage du codon UGA en sélénocystéine est sensiblement similaire, sur le principe, à celui des eubactéries.

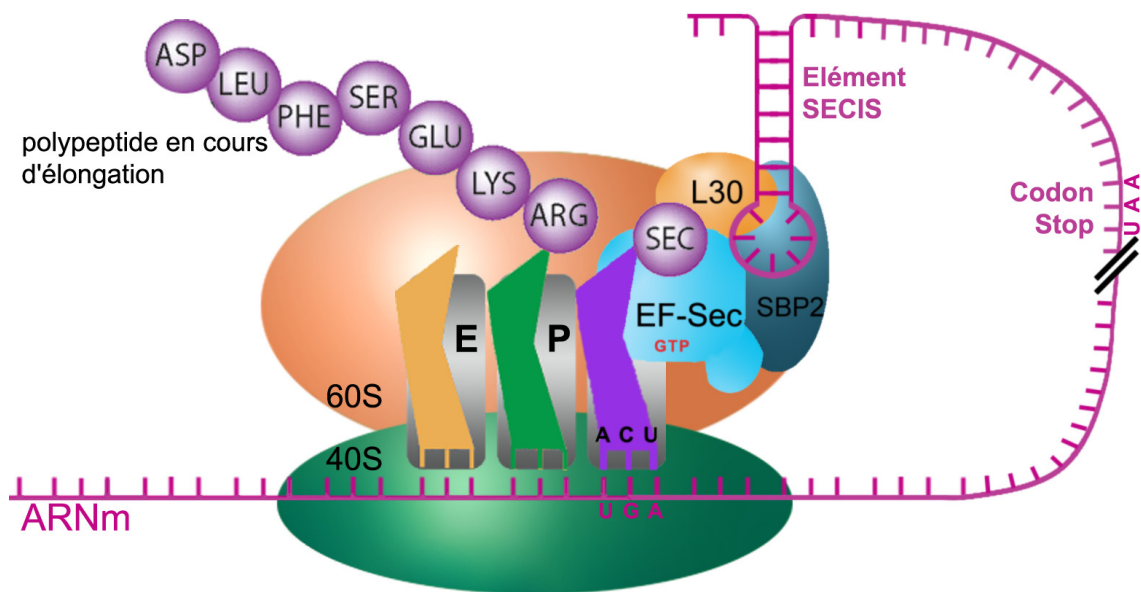
Les différences notables sont les suivantes :

- l'élément SECIS n'est pas présent immédiatement en aval du codon UGA à recoder, mais dans la région 3' non-codante de l'ARN<sub>m</sub>. Dans de rares cas, SECIS est même présente dans la partie 5' non-codante, en amont du gène.



**Figure 1.11** – Structures SECIS des eucaryotes<sup>[I.129]</sup> et des archaés.<sup>[I.130]</sup> Chez les eucaryotes, deux formes ont été caractérisées : leurs éléments conservés sont indiqués sur ce schéma. Chez les archaés, les éléments caractéristiques de la SECIS sont encadrés.

- La structure de l'élément SECIS possède d'autres caractéristiques, elles aussi conservées d'un gène à l'autre et permettant l'interaction avec une protéine spécifique. Ils sont représentés dans le schéma 1.11, page précédente. Chez les eucaryotes, ces éléments sont : un "quartet" composé de deux paires de base non Watson-Crick G-A et d'adénines, conservées dans la plupart des cas, sur la seconde boucle de la structure. Certaines formes de SECIS contiennent une troisième tige et une boucle apicale non-conservées. Chez les archaées, l'élément caractéristique de SECIS est une boucle conservée de quatre purines : deux adénines et une guanine, faisant face à une adénine. Une troisième hélice est également présente, composée de deux ou trois paires G-C et suivie d'une boucle apicale non-conservée.



**Figure 1.12** – Schéma représentant, chez les eucaryotes, la présentation au ribosome du Sec-ARN<sub>t</sub><sup>Sec</sup> par le facteur d'élongation spécifique EF-Sec, interagissant avec L30 et avec SBP2, qui reconnaissent la structure secondaire SECIS.

- Les plus grands changements interviennent au niveau des facteurs protéiques nécessaires à l'incorporation de la sélénocystéine. Alors que chez les eubactéries, SelB remplit le double rôle d'interaction avec SECIS et avec le Sec-ARN<sub>t</sub><sup>Sec</sup>, ces fonctions sont remplies chez les eucaryotes par la "SECIS Binding Protein" SBP2 et EF-Sec, respectivement.

- EF-Sec et son équivalent archaïque, nommé aSelB, sont des protéines homologues à la protéine SelB bactérienne, bien que plus courtes. Leurs extrémités C-terminales ne leur permettent pas d'interagir avec SECIS. Ces facteurs d'élongation spécifiques occupent le rôle de présenter le Sec-ARN<sub>t</sub><sup>Sec</sup> au ribosome.

- SBP2, dont l'équivalent archaïque n'a pas encore été trouvé, est la protéine qui remplit le rôle de la partie C-terminale de SelB : elle interagit avec SECIS et avec EF-Sec. C'est une

protéine qui comporte trois domaines, dont le troisième, en C-terminal, posséderait un motif L7Ae, connu pour interagir avec l'ARN.<sup>[I.131]</sup> Ce domaine interagirait avec SECIS et avec le ribosome, mais pas simultanément. La protéine ribosomale L30 possède également un motif L7Ae. Elle interagit également avec SECIS<sup>[I.132]</sup> et serait impliquée dans la réussite de l'incorporation de la sélénocystéine, par une interaction compétitive avec SECIS.<sup>[I.133]</sup>

La protéine SBP2 serait également responsable du transport hors du noyau de l'ARN<sub>m</sub> transcrit de sélénoprotéine, grâce à son interaction avec l'élément SECIS. Sa localisation dans le noyau ou le cytoplasme et son passage de l'un à l'autre seraient dépendants de son statut rédox, ce qui constituerait une forme de régulation de l'expression des sélénoprotéines.<sup>[I.134]</sup>

- Un facteur protéique supplémentaire à sa place dans le processus d'incorporation de la sélénocystéine : SECp43. Cette protéine interagit avec l'ARN<sub>t</sub><sup>Sec</sup> et son rôle est la méthylation de l'hydroxyl en 2' de l'uridine U34 (position "wobble") de l'ARN<sub>t</sub><sup>Sec</sup>, une fois que celui-ci est chargé avec une sélénocystéine.<sup>[I.135]</sup> Cette méthylation est la dernière étape dans la maturation du Sec-ARN<sub>t</sub><sup>Sec</sup> et affecte grandement la structure secondaire et tertiaire de l'ARN.<sup>[I.136]</sup> Les deux populations de Sec-ARN<sub>t</sub><sup>Sec</sup>, avant et après méthylation, n'auraient pas la même efficacité d'incorporation de la sélénocystéine. Une déficience en sélénium provoquerait la diminution de la population d'ARN<sub>t</sub><sup>Sec</sup> méthylé, ce qui a pour conséquence une diminution de l'expression des sélénoprotéines liées à la gestion du stress oxydatif, alors que les sélénoprotéines vitales impliquées dans le métabolisme énergétique, comme les thiorédoxine réductases, ne sont pas touchées par cette diminution.<sup>[I.136]</sup> Cela souligne un autre système de régulation de l'expression de certaines sélénoprotéines.

L'incorporation de la sélénocystéine lors de la synthèse de sélénoprotéine est donc beaucoup plus complexe chez les eucaryotes et chez les archaées (pour lesquelles des données sont encore manquantes) que chez les eubactéries, pour des raisons de régulation et des raisons fonctionnelles, comme le passage de la membrane nucléaire. Malgré les différences flagrantes que nous venons de voir, la similarité de ce principe d'incorporation conforte l'hypothèse d'une filiation commune entre ces mécanismes.

### 1.3.3 Le cycle de l'acide aminé libre

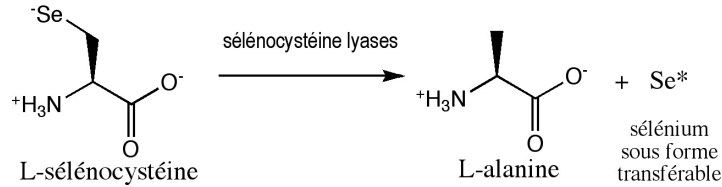
#### ou le recyclage du sélénium par la cellule

L'existence d'une protéine commence par la traduction de son ARN<sub>m</sub>, comme nous venons de le voir pour les sélénoprotéines, et se termine par sa dégradation et son recyclage. Les différents acides aminés résultants de cette dégradation sont utilisés dans des voies métaboliques qui leur sont propres, ou réutilisés dans la synthèse de nouvelles protéines.

Comme la sélénocystéine ne peut être réutilisée telle quelle, les organismes qui l'emploient ont mis en place des voies de recyclage efficace, pour éviter l'accumulation de ce métabolite délétère, mais aussi pour économiser cette ressource rare qu'est le sélénium.



Tous les organismes possèdent des enzymes capables de dissocier la cystéine en alanine et en soufre : les L-cystéine-lyases. Il en va de même pour la sélénocystéine, substrat des cystéine-lyases non-spécifiques ou de sélénocystéine-lyases.

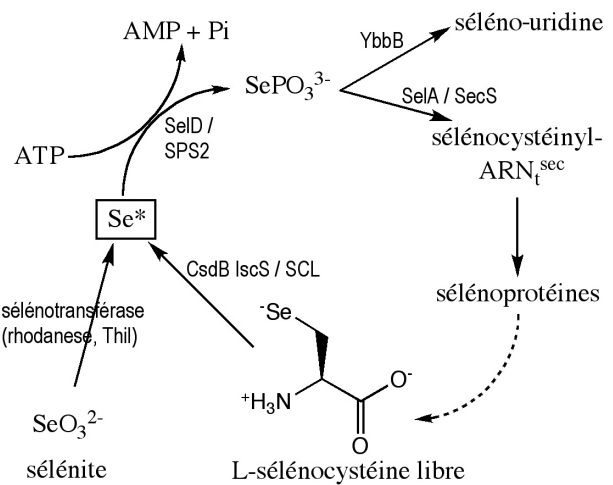


**Figure 1.13** – Réaction catalysée par les sélénocystéine-lyases.  $\text{Se}^*$  est, selon les cas, soit du  $\text{Se}^0$ , soit une espèce séléniée de transfert entre protéine.<sup>[1.137]</sup>

Des sélénocystéines-lyases ont été identifiées chez les eubactéries et les eucaryotes<sup>[1.138]</sup> où elles sont nommées SCL. Ce sont des protéines pyridoxal-phosphate-dépendantes faisant partie de la famille des protéines à domaine NifS.<sup>[1.137]</sup>

Chez *Escherichia coli*, trois protéines ayant une activité cystéine et sélénocystéine lyase ont été identifiées : CSD (ou CsdA), CsdB et IscS. Les structures tridimensionnelles de ces deux dernières sont disponibles en Annexe A.6, page 235. Parmi ces protéines, CsdB est 290 fois plus active sur la sélénocystéine que sur la cystéine<sup>[1.139]</sup> et pourrait être la SCL bactérienne.<sup>[1.140]</sup>

La formation d'un complexe protéique entre SelD et CsdB pour transférer le sélénium obtenu en vue de sa transformation ATP-dépendante en sélénophosphate avait été envisagée, mais ne pu être observée.<sup>[1.141]</sup> Un agent protéique supplémentaire inconnu est supposé effectuer cette opération de transfert, comme le fait la sulfur-transférase Thil dans le métabolisme du soufre. L'intérêt d'un tel agent de transfert est de limiter la concentration de l'élément libre en milieu cellulaire, et donc sa toxicité.



**Figure 1.14** – Cycle de recyclage de la sélénocystéine.<sup>[1.141]</sup>  $\text{Se}^*$  représente l'entité séléniée de transfert entre protéines, dont la nature est inconnue.

Bien que des implications dans les voies métaboliques du soufre aient été trouvées pour IscS (cette dernière étant impliquée dans la distribution de soufre pour la synthèse d'ARN<sub>t</sub> thiolés, de cofacteur molybdoptérine et de cluster Fer-Soufre), de récentes études suggèrent une implication importante de cette enzyme dans la voie de synthèse de la sélénocystéine, son absence entraînant une diminution de l'expression des sélénoprotéines.<sup>[1.142]</sup> Son importance dans le métabolisme du sélénium reste controversée.

Chez les eucaryotes, un acteur supplémentaire pourrait jouer un rôle dans le recyclage de la sélénocystéine : la sélénophosphate synthétase SPS1. Cette protéine, qui ne semble pas capable d'utiliser le sélénium minéral présent dans la cellule pour synthétiser du sélénophosphate,<sup>[I.143]</sup> utiliserait la sélénocystéine libre comme substrat de cette réaction, et combinerait ainsi les activités des lyases et SPS2.<sup>[I.144]</sup> Cette hypothèse est pourtant contredite par la non-suffisance de SPS1 pour permettre la synthèse de sélénoprotéines, lorsque SPS2 est absente.<sup>[I.122]</sup> L'énigme du rôle controversé de SPS1 n'a pas encore de solution, même si la récente découverte de sa présence chez des organismes incapables de synthétiser la sélénocystéine semble indiquer que sa fonction n'est pas en relation avec la synthèse de sélénoprotéines.<sup>[I.101]</sup>

Bien que tous les acteurs du recyclage de la sélénocystéine ne soient pas connus, ce système est efficace et permet aux organismes utilisant cet acide aminé de ne pas souffrir de sa toxicité.

### **La sélénocystéine, un acide aminé exceptionnel**

Nous venons de voir la synthèse, l'incorporation dans les protéines et le recyclage de la sélénocystéine. À chacune de ces étapes, cet acide aminé n'emprunte pas les voies communes aux 20 acides aminés canoniques, mais des systèmes spécifiques. Ainsi, il n'existe pas de SecRS, ni de codon défini spécifiquement pour l'incorporation de la sélénocystéine.

Quelle sont les raisons de ces caractéristiques exceptionnelles ? Peuvent-elles être expliquées par la toxicité de la sélénocystéine, par un possible ajout tardif dans le code génétique, ou par d'autres raisons encore insoupçonnées ? Sans apporter de réponses à ces interrogations, la comparaison de la sélénocystéine avec d'autres acides aminés particuliers pourrait apporter des éléments pour la compréhension de cette spécificité.

## **1.4 Autres acides aminés particuliers comparables**

La sélénocystéine n'est pas le seul acide aminé particulier, co-translationnellement incorporé dans les protéines. D'autres acides aminés partagent les caractères d'exception que sont l'ajout d'un nouvel ARN<sub>t</sub> au code génétique ou une synthèse indirecte de l'espèce finale aminoacyl-ARN<sub>t</sub>. Cette section 1.4 fait l'énumération des connaissances actuelles dans ce domaine, permettant de comparer ces autres acides aminés excentriques à la sélénocystéine.

### **1.4.1 L'asparagine et la glutamine**

Une des raisons pour lesquelles un organisme utiliserait une voie de synthèse indirecte de la brique de construction que constitue l'acide aminé estérifié à son ARN<sub>t</sub> est l'absence dans son génome d'une aminoacyl-ARN<sub>t</sub>-synthétase spécifique. Telle est la situation pour l'asparagine et la glutamine chez une majorité d'eubactéries et la quasi-totalité des archaées.<sup>[I.145]</sup>

Dans le cas de l'asparagyl-ARN<sub>t</sub><sup>Asn</sup>, par exemple, sa voie de synthèse indirecte implique d'abord une més-estérification d'un aspartate sur l'ARN<sub>t</sub><sup>Asn</sup> par une aspartyl-ARN<sub>t</sub><sup>Asp</sup> synthétase ayant une spécificité relâchée. On parle alors d'aspartyl-ARN<sub>t</sub> synthétase non-discriminante, ou ND-AspRS. L'Asp-ARN<sub>t</sub><sup>Asn</sup> subit ensuite une transamidation ARN<sub>t</sub>-dépendante catalysée par une amidotransférase (AdT) utilisant une asparagine ou une glutamine libre comme donneur d'ammonium.

Cette voie de synthèse a été découverte en 1968<sup>[I.146]</sup> pour la glutamine et en 1996<sup>[I.147]</sup> pour l'asparagine. L'utilisation répandue de cette voie de synthèse indirecte est devenue apparente grâce à la disponibilité d'un nombre de plus en plus important de génomes bactériens séquencés. On sait aujourd'hui qu'une majorité d'organites, d'eubactéries et la totalité des archaées séquencées ne possède pas de GlnRS et passe donc par la voie de la transamidation pour obtenir le Gln-ARN<sub>t</sub><sup>Gln</sup>. Le cas de l'asparagine est un peu différent : l'AsnRS est également manquante dans une majorité d'organites, chez les archaées sauf quelques exceptions, et chez beaucoup de bactéries Gram-positives, proteobactéries  $\alpha$ ,  $\beta$  et  $\epsilon$ , Chlamydiales, Aquificales et Thermotogales. Mais *Escherichia coli*, par exemple, possède les GlnRS et AsnRS, héritées d'un transfert horizontal de gènes en provenance d'eucaryotes.

Alors que chez certaines archaées, les deux voies de synthèse, directe et indirecte, peuvent cohabiter,<sup>[I.148]</sup> chez les eubactéries, seuls les 4 cas de non-redondance des voies de synthèse sont retrouvés parmi les organismes séquencés à ce jour : l'organisme possède soit les deux GlnRS et AsnRS, soit des ND-GluRS et ND-AspRS et l'AdT, soit l'ARN<sub>t</sub>-synthétase d'un des deux acides aminés et une voie de transamidation pour l'autre.

L'amidotransférase ARN<sub>t</sub>-dépendante la plus répandue se retrouve dans les trois domaines du vivant : il s'agit d'une enzyme hétérotrimérique encodée par les trois gènes organisés dans un opéron *gatCAB*. Elle n'est pas spécifique : elle catalyse à la fois la transamidation sur le Glu-ARN<sub>t</sub><sup>Gln</sup> et sur le Asp-ARN<sub>t</sub><sup>Asn</sup>.

Cependant, certaines archaées possèdent une AdT ARN<sub>t</sub>-dépendante différente, dimérique, encodée par les gènes *gatD* et *gatE*, et qui ne catalyse que la transamidation sur Glu-ARN<sub>t</sub><sup>Gln</sup>.

Par comparaison au cas de la sélénocystéine, une question vient immédiatement à l'esprit : existe-t-il un risque d'incorporation accidentelle des ARN<sub>t</sub> més-estérifiés Asp-ARN<sub>t</sub><sup>Asn</sup> ou Glu-ARN<sub>t</sub><sup>Gln</sup>, ou les organismes utilisant la transamidation possèdent-ils une sécurité contre ce risque ? En fait, il a été observé<sup>[I.148]</sup> que la capacité de liaison à EF-Tu n'est conférée à l'ARN<sub>t</sub><sup>Asn</sup> que par la conversion de l'aspartate estérifié en asparagine. Cette discrimination permet d'éviter tout risque, la situation étant sans doute similaire pour l'ARN<sub>t</sub><sup>Gln</sup>.

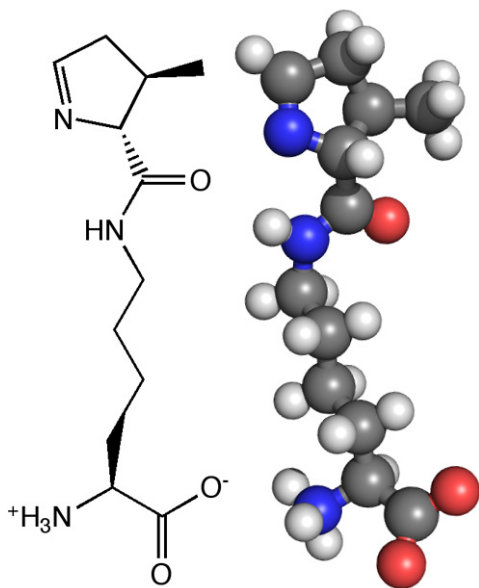
En ce qui concerne l'évolution des voies de synthèses des aminoacyl-ARN<sub>t</sub>, et en particulier l'émergence des aminoacyl-ARN<sub>t</sub> synthétases, le cas du couple aspartate / asparagine apporte des éléments intéressants. Il a été en effet démontré que, pour l'Asp-ARN<sub>t</sub><sup>Asp</sup> synthétase de *Thermococcus kodakaraensis*, la simple mutation d'un acide aminé dans la boucle de fixation de l'anticodon de l'ARN<sub>t</sub> fait passer l'enzyme de discriminante à non-

discriminante vis-à-vis de  $\text{ARN}_t^{\text{Asn}}$ .<sup>[I.149]</sup> Cela implique qu'une mutation inverse aurait pu transformer une enzyme non-discriminante en enzyme spécifique, ce qui ouvre le chemin d'une hypothèse sur l'évolution divergente des AspRS et ND-AspRS et sur l'émergence de la voie d'asparagylation directe. Il est admis que les AsnRS et GlnRS sont les aaRS dont l'émergence est la plus récente. Il s'agit d'un événement ayant eu lieu après le stade du dernier ancêtre commun entre les trois domaines du vivant, appelé LUCA (Last Universal Commun Ancestor).<sup>[I.150]</sup> Cela nous rappelle que les aminoacyl- $\text{ARN}_t$ -synthétases ne sont que des produits d'adaptation au code génétique pré-existant, incarné par les  $\text{ARN}_t$  qui ont, eux, établi le catalogue des codons.

### 1.4.2 La pyrrolysine, le 22<sup>ème</sup> Acide Aminé

Voici un acide aminé qui a été rajouté tardivement à notre connaissance du code génétique, et cela à cause de la restriction de son existence actuelle à seulement quelques rares organismes. Sa découverte date de 2002,<sup>[I.151]</sup> lors de la résolution de la structure tridimensionnelle d'une monométhylamine méthyltransférase d'archae méthanogène.

Cette N6-[(2R,3R)-3-méthyl-3,4-dihydro-2H-pyrrol-2-ylcarbonyl]-L-lysine a la particu-



larité de se situer dans la séquence d'acides aminés de cette protéine à une position correspondant à un codon Stop UAG, ou ambre, du gène correspondant. L'incorporation de la pyrrolysine lors de la traduction protéique implique un  $\text{ARN}_t^{\text{Pyl}}$  spécifique, possédant un anticodon CUA, "Ambre-suppresseur".

Cette particularité d'incorporation co-translationnelle sur un codon Stop, commune avec la sélénocystéine, a guidée la recherche de points communs entre le 21<sup>ème</sup> et le 22<sup>ème</sup> acide aminé. Pourtant, ce nouvel acide aminé encodé génétiquement, symbolisé par les codes à trois lettres Pyl et à une lettre O, n'a pas beaucoup de choses en commun avec la sélénocystéine.

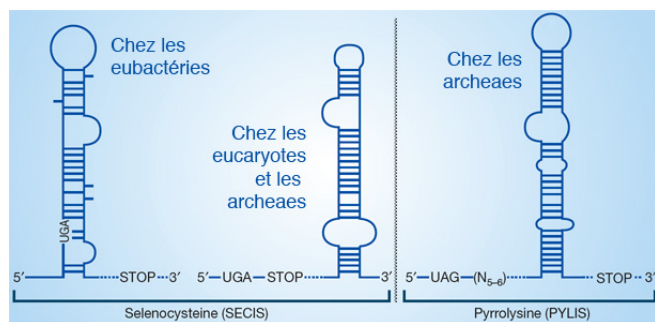
**Figure 1.15** – Formule de la Pyrrolysine.

Le modèle actuellement plébiscité, basé sur des données expérimentales *in vitro* et *in vivo*, est celui d'une incorporation co-translationnelle simple, possible grâce à l'ajout pur et simple de la pyrrolysine au code génétique chez les organismes concernés. Ainsi, chez ces organismes, la réaffectation du codon UAG pour la pyrrolysine a été possible grâce à l'émergence de l' $\text{ARN}_t^{\text{Pyl}}$ , codé par le gène *pylT*, et de la PylRS, codée par le gène *pylS*. Chez ses organismes, la distinction entre les sens "Stop" et "incorporation de pyrrolysine", soit n'est plus nécessaire car le codon UAG a été recodé pour l'insertion de la pyrrolysine, soit suit un mécanisme actuellement indéterminé, pour lequel beaucoup d'hypothèses sont postulées. Reprenons les données qui ont permis d'arriver à ces résultats.

Un résumé des connaissances actuelles sur le sujet serait le suivant :

- La pyrrolysine ne concerne que 6 organismes méthanogènes et un endosymbiote, parmi les organismes dont le génome est séquencé à ce jour : l'archae *Methanococcoides burtonii*,<sup>[I.152]</sup> les quatre représentants de la famille d'archaes *Methanosarcina genera* que sont *M. barkeri*, *M. mazei*, *M. acetivorans* et *M. thermophila*, l'eubactérie *Desulfitobacterium hafniense*<sup>[I.153]</sup> et l'endosymbiote  $\delta$ -protéobactérien  $\delta$ 1 d'*Olavius algarvensis*.<sup>[I.154]</sup>
- Chez ces organismes sont retrouvées 3 familles de méthyltransférases initiant la méthanogenèse à partir de la réduction de méthylamines,<sup>[I.155]</sup> et ne possédant que peu de similarité de séquences entre elles : les monométhylamine-, diméthylamine- et triméthylamine-transférases (ou MMA-, DMA- et TMA-méthyltransférases) codées respectivement par les gènes *mtmB*, *mtbB* et *mttB*. Les gènes des méthyltransférases possèdent un codon ambre dans leur cadre de lecture. Cette découverte effectuée en 2000<sup>[I.155]</sup> fut suivie par la confirmation, en 2001,<sup>[I.156]</sup> de l'utilisation de ce codon comme un codon sens, conduisant à l'incorporation co-translationnelle d'un acide aminé, qui, d'après des analyses par dégradation d'Edman et spectrométrie de masse, pouvait être une lysine ou un dérivé.
- En 2002, la structure de la pyrrolysine fut révélée grâce aux cartes de densité électronique obtenues par diffraction aux rayons X de cristaux d'une MMA-aminotransférase de *M. barkeri*.<sup>[I.151]</sup> Au même moment furent découverts<sup>[I.157]</sup> les gènes *pylT* et *pylS* codant respectivement pour l'ARN<sub>t</sub><sup>Pyl</sup> et la PylRS.
- Comme la pyrrolysine est incorporée par le recodage d'un codon Stop, tout comme la sélénocystéine, l'existence d'un guidage de l'incorporation par une séquence signal a été spéculée dès 2002. Une structure secondaire de l'ARN<sub>m</sub> peut en effet être prédite à la suite du codon UAG. Par homologie à la séquence signal SECIS, ce possible message de l'ARN<sub>m</sub> a été nommé PYLIS pour "PYrroLysine Insertion Sequence".

À partir de 2005, la nécessité d'une PYLIS devient controversée<sup>[I.159]</sup> suite à la découverte de nouveaux éléments. D'un côté, il a été observé une conservation de la séquence de la partie PYLIS dans les gènes *mtmB*,<sup>[I.160]</sup> ce qui confirme l'importance de cette zone dans la séquence. D'un autre côté, chez *E. coli*, un gène comme *lacZ* contenant par mutation UAG dans sa séquence conduit tout de même à l'expression d'une protéine complète et active, lorsque l'organisme co-exprime les gènes *pylTS* et qu'un dérivé de la pyrrolysine est présent dans le milieu.<sup>[I.161]</sup> Cela tant à montrer qu'aucune séquence signal n'est nécessaire pour le recodage d'un codon UAG. Encore récemment, deux études sont arrivées à



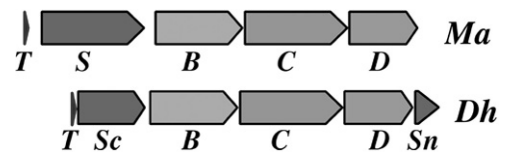
**Figure 1.16** – Comparaison des SECIS et PYLIS.<sup>[I.158]</sup>

des conclusions contraires au sujet de la conséquence de l'absence ou de la présence de la PYLIS : O. Namy<sup>[1.162]</sup> montre que cette absence ou présence n'a aucune conséquence sur le niveau d'incorporation réussie de Pyl, chez *E. coli*, alors que D. G. Longstaff<sup>[1.163]</sup> montre que la présence d'une PYLIS améliore le rapport lecture du codon UAG / terminaison, tout en concédant qu'un gène exogène, contenant un codon UAG et incorporé chez une archae méthanogène, bénéficie tout de même d'une efficacité d'incorporation de la pyrrolysine de 20%, sans PYLIS.

Des études bio-informatiques apportent d'intéressants éléments complémentaires grâce à l'analyse des génomes des organismes utilisant la pyrrolysine.<sup>[1.164]</sup> Deux des principales conclusions sont que les séquences en aval du codon UGA, où devrait se trouver la PYLIS, sont dissemblables dans les gènes *mtmB*, *mtbB* et *mttB*, et que le codon UGA n'est pas un codon Stop typique chez les archaees utilisant la pyrrolysine. Cette dernière affirmation vient du calcul de l'utilisation des différents codons Stop chez ces archaees : les gènes prédits se terminant par UAG (dont les méthyltransférases incorrectement annotées) comptent pour moins de 5% du total des gènes chez *M. mazei*, *M. barkeri* et *M. acetivorans*, contre ~46% pour UGA et ~49% pour UAA. De plus, les codons UAG qui sont supposés avoir le non-sens Stop, sont quasiment toujours suivis d'un autre codon Stop de nature différente,<sup>[1.165]</sup> sans doute pour minimiser dans ces cas les conséquences d'un possible recodage non-voulu.

Étonnamment, la situation est différente chez *D. hafniense*, où le pourcentage de gènes prédits se terminant par UAG atteint 22,5%, ce qui est similaire à la proportion de ceux se terminant par UGA (28,4%). Cette utilisation étendue du codon UAG chez l'eubactérie suggère une dualité de son rôle dans la synthèse protéique, sans que l'on sache aujourd'hui ce qui permet à l'organisme de lever cette ambiguïté lors de la traduction.

- D'autres données permettent de comprendre le fonctionnement de la synthèse et de l'incorporation de la pyrrolysine dans ces organismes. Les gènes *pylT* et *pylS* se retrouvent sous forme d'opéron, avec trois autres gènes, *pylBCD*,<sup>[1.166]</sup> nécessaires pour la biosynthèse de la pyrrolysine à partir de métabolites plus simples, comme la D-ornithine.<sup>[1.162]</sup> L'ordre des gènes *pylTSBCD* est maintenu chez les archaees méthanogènes, mais également chez l'eubactérie *D. hafniense*, dont la seule particularité réside en la séparation du gène *pylS* en deux parties, correspondant aux parties C- et N-terminales de la PylRS. Il a été montré que le transfert de l'opéron à un organisme "naïf" lui permet l'acquisition des compétences de biosynthèse et d'incorporation de la pyrrolysine au niveau de codons UAG.<sup>[1.166]</sup> Cela appuie la vraisemblance de l'acquisition de ces compétences par *D. hafniense* grâce à un transfert horizontal de gènes archaees.



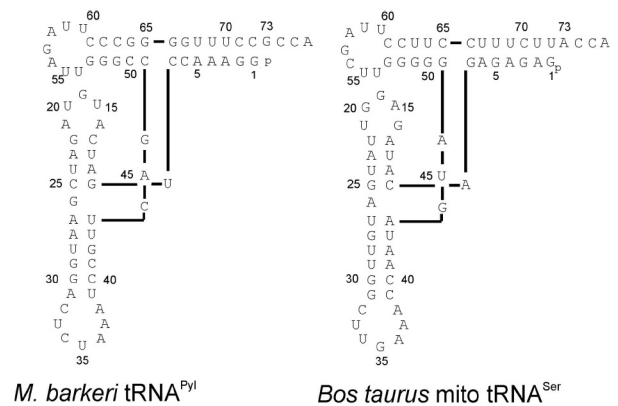
**Figure 1.17** – Structure en opéron des gènes *pylTSBCD* chez *M. acetivorans* et *D. hafniense*.<sup>[1.166]</sup>

- Le produit du gène *pylT*, l'ARN<sub>t</sub><sup>Pyl</sup>, est reconnu par EF-Tu comme n'importe quel ARN<sub>t</sub> canonique<sup>[1.167]</sup> : il s'agit d'un "ambre-suppresseur" ne nécessitant pas d'autre facteur protéique pour être utilisé lors de la traduction.<sup>[1.168]</sup>

- Les seuls gènes *pylTS* suffisent à induire une incorporation de la pyrrolysine aux codons UAG chez *Escherichia coli*, lorsque le milieu contient cet acide aminé, et à étendre ainsi son code génétique.<sup>[I.168]</sup> Le couple  $\text{ARN}_t^{\text{Pyl}}$  et PylRS se comporte donc comme les couples  $\text{ARN}_t/\text{ARN}_t$  synthétase classiques.<sup>[I.169]</sup>

- Pourtant à sa découverte, le gène *pylS* était supposé coder pour une synthétase liant une lysine à l' $\text{ARN}_t^{\text{Pyl}}$ .<sup>[I.157]</sup> Cette més-esterification hypothétique aurait nécessité une réaction de transformation de l'acide aminé  $\text{ARN}_t$ -dépendante, comme dans le cas de la sélénocystéine. On sait aujourd'hui que la PylRS est capable d'esterifier directement la pyrrolysine à l' $\text{ARN}_t^{\text{Pyl}}$ ,<sup>[I.168]</sup> mais aussi la lysine et des dérivés comme la N- $\epsilon$ -D-prolyl-L-lysine (D-prolyl-lysine) ou la N- $\epsilon$ -cyclopentylloxycarbonyl-L-lysine (Cyc), *in vitro*.<sup>[I.161]</sup>

Quant au fonctionnement de l'enzyme, des études montrent que le domaine N-terminal est indispensable pour l'activité PylRS *in vivo*, mais est facultatif *in vitro*,<sup>[I.170]</sup> et que l'interaction avec l' $\text{ARN}_t^{\text{Pyl}}$  ne s'effectue pas par reconnaissance de l'anticodon, mais par la reconnaissance de caractères d'identités classiques des  $\text{ARN}_t$ , comme la base discriminante G73, la première paire de bases de la tige acceptrice, la paire G5-C63 et les nucléotides U33 et A37 épaulant l'anticodon. La structure de l' $\text{ARN}_t^{\text{Pyl}}$  est très classique, et peut même être comparée à l' $\text{ARN}_t^{\text{Ser}}$  mitochondrial bovin, très similaire.<sup>[I.169]</sup> (voir figure 1.18)



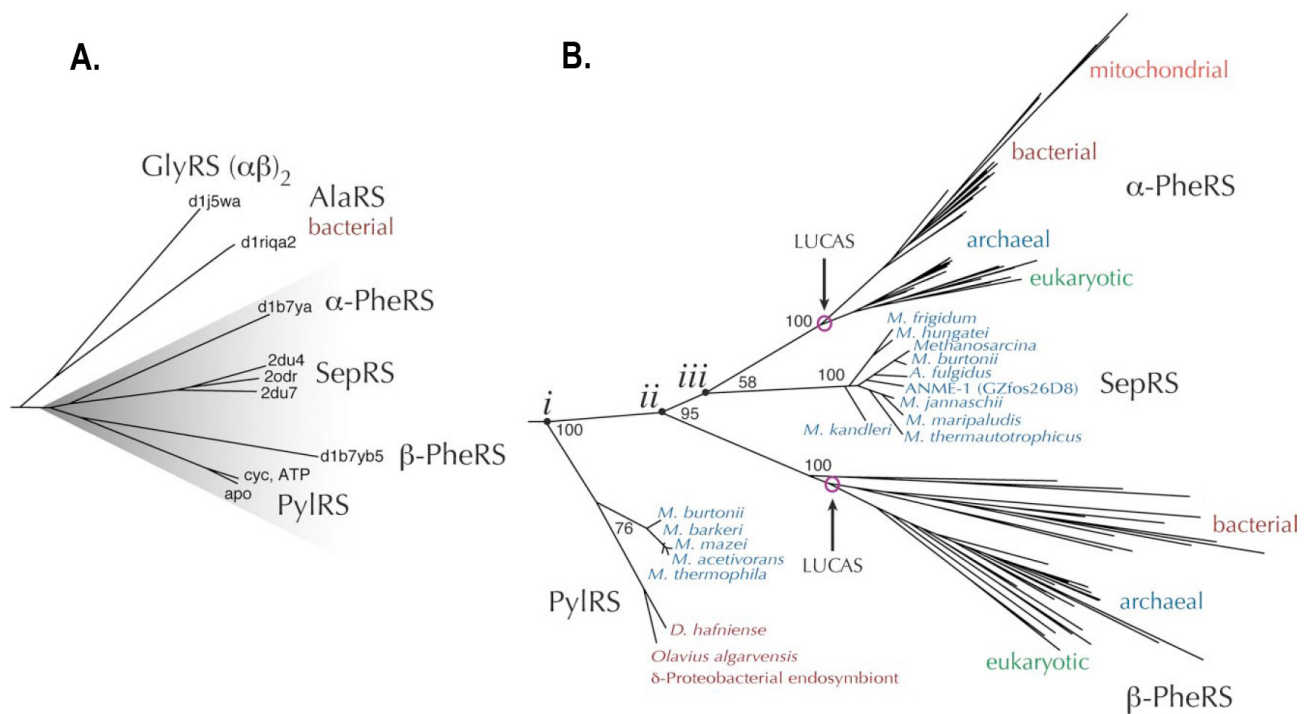
**Figure 1.18** – Comparaison des structures secondaires d' $\text{ARN}_t^{\text{Pyl}}$  archaïque et d' $\text{ARN}_t^{\text{Ser}}$  mitochondrial bovin.<sup>[I.169]</sup>

Il a également été observé que la reconnaissance de l'acide aminé libre à esterifier par PylRS se fait grâce à une poche hydrophobe. Cette dernière donnée est disponible grâce à la résolution de la structure de la PylRS de *M. mazei* en 2007.<sup>[I.171]</sup>

La PylRS paraît appartenir à la classe II des aminoacyl- $\text{ARN}_t$ -synthétases, par similitude structurale avec les glycyl- $\text{ARN}_t$  synthétase, alanyl- $\text{ARN}_t$  synthétase, phosphosérine- $\text{ARN}_t$  synthétase (SepRS) et les unités divergentes  $\alpha$ - et  $\beta$ - de la phenylalanyl- $\text{ARN}_t$  synthétase.

Bien que la similarité de structure tertiaire entre la PylRS et les autres aaRS de la classe II soit très importante, les alignements de séquences ne permettent pas d'obtenir des relations consensuelles entre ces enzymes. C'est donc par combinaison des comparaisons de structures et de séquence que l'arbre phylogénétique en figure 1.19 a pu être construit. Comme les enzymes AlaRS et SepRS sont les seules aaRS homotétramériques connues et que PheRS est un hétérotétramère dont les deux sous-unités restent fortement apparentées, il est possible que PylRS soit également un tétramère en solution.

D'un point de vue évolutif, il est à noter que la PylRS et la SepRS (voir section 1.4.3) ne sont représentées aujourd'hui que dans un seul domaine du vivant : celui des archaées (un transfert horizontal de gènes archaées expliquant la présence de PylRS chez deux eubactéries). Pourtant, la phylogénie de la PylRS l'affilie à la PheRS primitive, avant sa divergence en unités  $\alpha$  et  $\beta$ , antérieure au stade du dernier ancêtre commun entre les domaines du vivant, LUCAS.<sup>[I.171]</sup> (voir schéma 1.19) Cela permet de penser que la PylRS est apparue après la PheRS dans le paysage des enzymes responsables du maintien du code génétique. Le fait que cette extension du code génétique par la pyrrolysine ne soit retrouvée que dans un nombre de familles restreint d'organismes laisse supposer, soit la perte précoce de cette capacité par la plupart des organismes primitifs, lorsque leur métabolisme énergétique n'eut plus le besoin de s'appuyer sur la réduction des méthylamines, soit l'émergence restreinte de la PylRS, chez certaines archaées, à partir d'une copie de la PheRS primitive présente dans leur génome depuis le stade LUCAS.



**Figure 1.19** – Arbres Phylogéniques des ARN<sub>t</sub> synthétases de la sous-classe IIc.<sup>[I.171]</sup>

(A) Phylogénie structurale de la classe IIc aux côtés des sous-classes IIa et IIb.

(B) Phylogénie structurale et par alignement de séquence de la sous-classe IIc.

### 1.4.3 La cystéine chez certaines Archaes

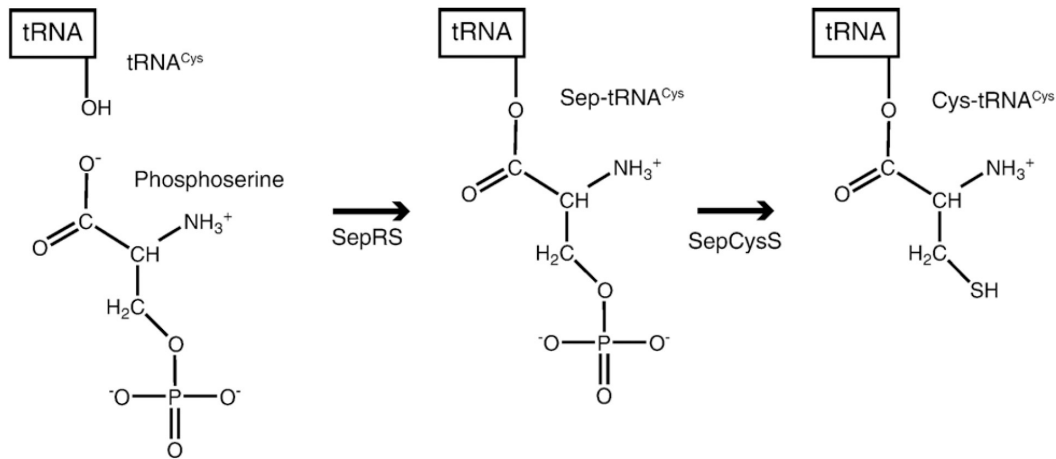
Le paragraphe précédent introduit l'existence d'une aminoacyl-ARN<sub>t</sub> synthétase non apparentée à un acide aminé canonique : la SepRS. Cette enzyme estérifie une phosphosérine sur l'ARN<sub>t</sub><sup>Cys</sup> chez certaines archaées méthanogènes.

Il s'agit de la première étape de la biosynthèse du cystéinyl-ARN<sub>t</sub><sup>Cys</sup> chez ces organismes, qui ne possèdent exceptionnellement pas de CysRS.<sup>[I.172]</sup>



La seconde étape de cette biosynthèse est la conversion  $\text{ARN}_t$ -dépendante du Sep- $\text{ARN}_t^{\text{Cys}}$  en  $\text{Cys-ARN}_t^{\text{Cys}}$  grâce à la phosphosérine-cystéine synthase  $\text{ARN}_t$ -dépendante (SepCysS), à l'image de la biosynthèse eucaryote de la sélénocystéine qui passe par un intermédiaire Sep- $\text{ARN}_t^{\text{Sec}}$ . Compte-tenu de ce parallèle, nous allons nous intéresser en détails aux connaissances actuellement disponibles sur cette voie de biosynthèse de la cystéine.

La découverte de cette voie de synthèse originale du  $\text{Cys-ARN}_t^{\text{Cys}}$  découle de la résolution d'une énigme mise à jour en 1999<sup>[I.173]</sup> : comme les génomes de certaines archaées ne semblent pas contenir de gène *cysS* codant pour la cystéinyl- $\text{ARN}_t$  synthétase, quelle est leur voie de synthèse du  $\text{Cys-ARN}_t^{\text{Cys}}$ , essentiel pour la synthèse protéique ?



**Figure 1.20** – Voie biosynthétique du  $\text{Cys-ARN}_t^{\text{Cys}}$  chez certaines archaées méthanogènes.<sup>[I.174]</sup> Cette voie peut même être la seule voie de synthèse de la cystéine chez certains organismes.

Aujourd'hui sont répertoriées plus d'une dizaine d'archaées utilisant la voie indirecte de synthèse de  $\text{Cys-ARN}_t^{\text{Cys}}$  : *Methanococcus jannaschii*,<sup>[I.173]</sup> *Methanobacterium thermoautotrophicum*,<sup>[I.173]</sup> *Methanopyrus kandleri*,<sup>[I.175]</sup> *Methanomicrobia frigidum*,<sup>[I.176]</sup> *Methanomicrobia hungatei*,<sup>[I.176]</sup> *Archaeoglobus fulgidus*,<sup>[I.176]</sup> *Methanococcoides burtonii*<sup>[I.176]</sup> et les représentants de la famille d'archaées *Methanosarcinaceae*,<sup>[I.176]</sup> que l'on retrouve dans le cas de la pyrrolysine (cf § 1.4.2, page 51). Il faut ajouter à cette liste *Methanococcus maripaludis*,<sup>[I.177]</sup> qui fait partie des organismes possédant les deux voies de synthèse du cystéinyl- $\text{ARN}_t^{\text{Cys}}$ , à la fois directe (CysRS) et indirecte (SepRS/SepCysS). Cette redondance montre que la voie de synthèse indirecte du  $\text{Cys-ARN}_t^{\text{Cys}}$  est trop importante dans leur métabolisme pour que ces organismes puissent s'en passer, alors que l'espèce est viable lors de la délétion du gène *cysS* de la CysRS.<sup>[I.177]</sup>

Comment cette voie de synthèse indirecte du  $\text{Cys-ARN}_t^{\text{Cys}}$  a-t-elle été découverte ?

Cela débuta, par deux fois, avec des résultats faussement-positifs. En 2000,<sup>[I.178]</sup> l'activité cystéinyl-ARN<sub>t</sub> synthétase fut trouvée dans une fraction de purification contenant la prolyl-ARN<sub>t</sub> synthétase, qui fut alors suspectée d'avoir à la fois l'activité ProRS et CysRS. Cela aurait bouleversé le dogme de fonctionnement substrat-spécifique des aminoacyl-ARN<sub>t</sub> synthétases faisant encore consensus aujourd'hui.

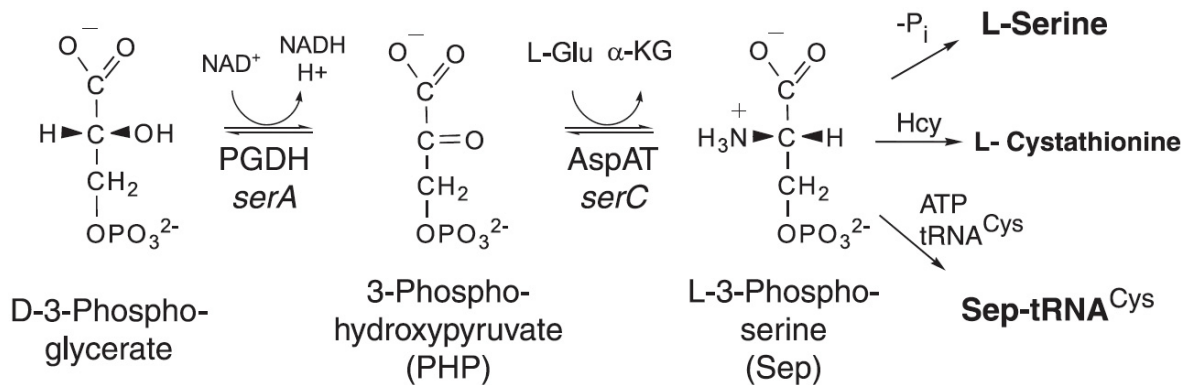
Puis en 2001,<sup>[I.179]</sup> l'activité CysRS fut imputée à l'enzyme MJ1477 de *M. jannaschii*, qui était décrite comme une CysRS atypique, ne possédant pas les motifs caractéristiques retrouvés chez tous les autres aaRS connues jusque-là.

Mais ces deux mini-révolutions s'avèrent découler d'une expérience de même type soulignant une activité CysRS *in vitro* inexistante, faille révélée en 2004.<sup>[I.175]</sup> De plus, le caractère anaérobie des organismes n'était pas pris en compte lors des purifications protéiques,<sup>[I.176]</sup> ce qui aurait conduit à des résultats trompeurs.

Ce retour à la case départ<sup>[I.180]</sup> ne dura pas longtemps puisque la voie de synthèse indirecte fut mise à jour en 2005.<sup>[I.176]</sup> L'activité CysRS fut trouvée dans une fraction de purification protéique contenant en majorité deux protéines, MJ1660 et MJ1678, conservées chez les archaées déficiente en CysRS.

MJ1660 apparaît être une enzyme homologue aux aaRS de classe II comme PheRS et ayant une activité ARN<sub>t</sub>-synthétase spécifique à la phosphosérine (Sep). Cette SepRS, codée par le gène *sepS*, est la seule aaRS connue estérifiant un acide aminé non-déterminé à la traduction, sur un ARN<sub>t</sub>.

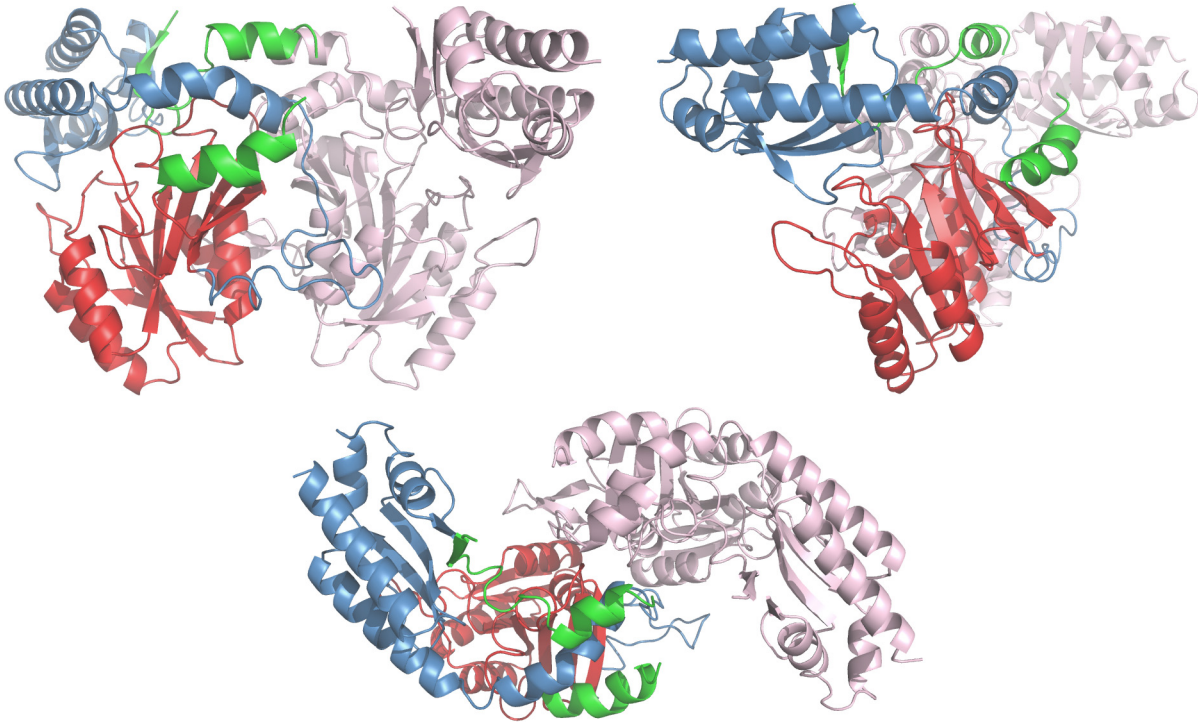
MJ1678 est une enzyme pyridoxal-phosphate-dépendante ayant une activité Sep-ARN<sub>t</sub><sup>Cys</sup> Cys-ARN<sub>t</sub><sup>Cys</sup> synthase. Cette SepCysS, codée par le gène *pseS*, est la seule enzyme connue à ce jour ayant une activité ARN<sub>t</sub>-dépendante comparable à celle des sélénocystéine synthases, ce qui en fait une enzyme de grand intérêt pour notre étude.



**Figure 1.21** – Biosynthèse et utilisation de la phosphosérine chez les *Methanococcales*.<sup>[I.181]</sup> Oxydation par PGDH, puis transamination par AspAT. (α-KG : α-kétoglutarate, Hcy : homocystéine)

La phosphosérine est donc le précurseur de la cystéine, sous sa forme cystéinyl-ARN<sub>t</sub><sup>Cys</sup>, et hérite à ce titre d'une aminoacyl-ARN<sub>t</sub> synthétase propre. Les études sur le métabolisme de la phosphosérine, chez les archaées citées précédemment, ont permis de découvrir que ces organismes donnent beaucoup d'importance à ce métabolite.(voir figure 1.21).

Ainsi la phosphosérine est le précurseur observé pour la synthèse de la cystéine et de l'homocystéine libres chez certaines de ces archaées,<sup>[1.182]</sup> mais aussi de la sérine et de la L-cystathionine.<sup>[1.181]</sup> Ce métabolite est principal chez les archaées synthétisant indirectement le Cys-ARN<sub>t</sub><sup>Cys</sup>, puisqu'il a été montré que certaines archaées, comme *M. jannaschii* et *M. maripaludis*, ne possèdent pas de voie de biosynthèse de la cystéine libre, et se reposent entièrement sur la voie ARN<sub>t</sub> dépendante pour biosynthétiser la cystéine.<sup>[1.176]</sup>



**Figure 1.22** – Structure de la SepCysS dimérique, accessible dans la PDB sous la référence 2E7J.<sup>[1.174]</sup> Un des deux monomères est coloré de façon suivante : partie N-terminale en vert, corps caractéristique en rouge, partie C-terminale en bleu.

Peu de temps après les études biochimiques de caractérisation de l'activité de SepRS et SepCysS, les structures de ces enzymes ont été résolues par deux équipes différentes, en 2007 : S. Kamtekar *et al.* fournirent la structure de la SepRS de *Methanococcus jannaschii*,<sup>[1.183]</sup> puis R. Fukunaga et S. Yokoyama publièrent la structure de la SepRS de *Archaeoglobus fulgidus* complexée à l'ARN<sub>t</sub><sup>Cys</sup> et à une phosphosérine,<sup>[1.184]</sup> suivie de la structure de SepCysS, toujours d'*Archaeoglobus fulgidus*.<sup>[1.174]</sup> Cette dernière est présentée figure 1.22.

Bien que la structure du complexe n'ait pu être obtenue, R. Fukunaga et S. Yokoyama proposent un modèle pour l'interaction entre l'ARN<sub>t</sub><sup>Cys</sup> et la Sep-ARN<sub>t</sub><sup>Cys</sup> Cys-ARN<sub>t</sub><sup>Cys</sup> synthase, visible figure 1.23, page suivante. Ce modèle est à mettre en comparaison avec le modèle d'interaction entre la sélénocystéine synthase eucaryote et l'ARN<sub>t</sub><sup>Sec</sup>, que nous allons voir dans la section qui lui est consacrée, page 102.

L'idée a en effet été émise<sup>[I.176]</sup> qu'étant donnée la similarité entre cette voie de synthèse de la cystéine chez certaines archaées et celle de la sélénocystéine chez les eucaryotes, une de ces deux voies aurait pu émerger sur les bases de l'autre. Ces deux voies passent en effet par une conversion du Sep-ARN<sub>t</sub> en Cys-ARN<sub>t</sub> ou en Sec-ARN<sub>t</sub>, en présence respectivement d'un donneur de soufre ou de sélénium approprié ce qui peut être considéré comme chimiquement analogue.

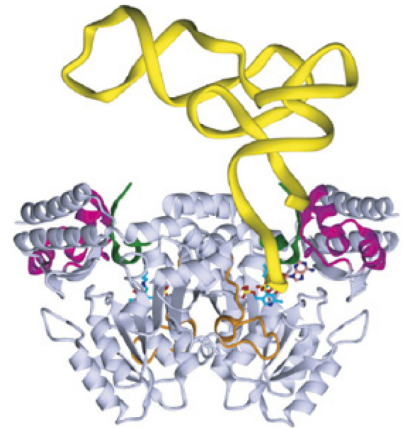
Cet intéressante hypothèse demande de se pencher plus spécialement sur les informations phylogénétiques que peuvent apporter des structures de la SepRS et de la SepCysS.

Comme nous l'avons vu pour PylRS à la section précédente (voir l'arbre phylogénique figure 1.19, page 55), SepRS et PylRS font toutes deux partie de la classe IIc des aaRS, et sont proches de la PheRS.<sup>[I.171]</sup> Il s'agit d'homotétramères et SepRS est structurellement plus proche de la sous-unité  $\alpha$ -PheRS que de la sous-unité  $\beta$ -PheRS.<sup>[I.183]</sup> Ces études conduisent aux mêmes conclusions : SepRS et PylRS étaient déjà présentes dans la communauté ancestrale, avant LUCAS, même si elles ne se retrouvent aujourd'hui quasiment que chez des archaées méthanogènes.

La SepCysS est quant à elle reliée phylogénétiquement aux cystéine-désulfurases des groupes I (les cystéine-sulfinate-désulfinate-ase CSB) et II (les protéines NiFS et IscS).<sup>[I.185]</sup> Toutes ces enzymes sont pyridoxal-phosphate-dépendantes, et catalysent un transfert de soufre. Des conservations de séquences entre ces enzymes laissent penser qu'elles proviennent d'un gène ancestral commun, et qu'elles peuvent partager un mécanisme proche.<sup>[I.185]</sup>

Le mécanisme de synthèse indirecte de la cystéine, présent chez ces archaées, soulève par son existence même des questions sur l'évolution des aaRS à partir du code génétique pré-existant. Si le code génétique universel était déjà établi lors du stade LUCAS, les voies d'acylation des ARN<sub>t</sub> ne l'étaient pas toutes. Une hypothèse pourrait être la suivante : les voies d'acylation par des aaRS, communes aujourd'hui aux trois domaines du vivant, avaient déjà émergées au stade LUCAS, alors que les ARN<sub>t</sub> qui bénéficient de voies d'acylation différentes selon les espèces témoignent d'une émergence plus tardive de voies utilisant les aaRS, comme dans les cas de la glutamine, l'asparagine, la cystéine et la lysine (qui est acylée sur son ARN<sub>t</sub><sup>Lys</sup> par des aaRS de classes différentes selon l'organisme).

Il est même envisageable que les voies alternatives de synthèse d'aminoacyl-ARN<sub>t</sub> exposées dans ce chapitre soient des reliques d'un fonctionnement d'acylation plus répandu au stade LUCAS. Cela soutient également l'hypothèse situant l'asparagine, la glutamine, et finalement la cystéine comme les derniers ajouts au code génétique, lors de son émergence parmi la communauté ancestrale.<sup>[I.176]</sup>



**Figure 1.23** – Modèle d'interaction entre SepCysS et l'ARN<sub>t</sub><sup>Cys</sup>.<sup>[I.174]</sup>

Ces notions ajoutent à l'étude de la sélénocystéine un attrait supplémentaire : cette voie de synthèse indirecte d'un acide aminé sur son  $\text{ARN}_t$ , présente parmi les trois domaines du vivant, est, elle aussi, une relique de l'époque ancestrale qui a vu la divergence des organismes vivants actuellement. La sélénocystéine est, à ce titre et dans l'état actuel des connaissances, plus proche de la cystéine archae que de la pyrrolysine.

## 1.5 L'avantage catalytique de la sélénocystéine

Nous venons de voir les cas d'acides aminés dont les ajouts au code génétique sont suffisamment "récents" pour que les voies de synthèse de leurs aminacyl- $\text{ARN}_t$  aient conservé la trace de voies ancestrales. Pourtant, dans tous les cas cités, une aminoacyl- $\text{ARN}_t$  synthétase spécifique a émergé et est maintenant présente chez une grande majorité d'organismes.

Pourquoi n'est-ce pas le cas pour la synthèse de la sélénocystéine ?

Les organismes utilisant cet acide aminé conservent une voie de synthèse qui non seulement n'est pas efficace mais qui en plus est très coûteuse : la synthèse et l'intégration de la sélénocystéine nécessite trois facteurs protéiques particuliers chez les eubactéries, et plus de six chez les eucaryotes, poids auquel il faut ajouter l'obligation de maintenir une structure particulière de message génétique pour conserver l'ambiguïté du sens du codon UGA. Si la toxicité de la sélénocystéine permet d'expliquer le fait que la synthèse du  $\text{Sec-ARN}_t^{\text{Sec}}$  n'a pu être transformée en une voie plus simple et classique, impliquant un codon propre, une SecRS et la présence de l'acide aminé libre dans la cellule, cela n'explique pas pourquoi tous les organismes n'ont pas abandonné cet acide aminé, comme la plupart ont pu le faire au sujet de la pyrrolysine (selon une des hypothèses sur l'émergence de celle-ci, avant LUCAS, voir § 1.4.2, page 55).

Manifestement, ces organismes considèrent la sélénocystéine suffisamment précieuse pour avoir permis l'émergence de cette voie de synthèse particulière, et pour entretenir sa conservation. Ce fait est d'autant plus surprenant que la plupart des sélénoprotéines possèdent des homologues de même fonction, où une cystéine se situe à la place de la sélénocystéine catalytique. Si la présence d'un atome de sélénium à la place d'un atome de soufre est si importante pour les organismes utilisant la sélénocystéine et non la cystéine, la raison doit se trouver au niveau des différences chimiques entre les deux acides aminés.

### 1.5.1 Propriétés chimiques de la sélénocystéine

Le soufre et le sélénium sont très proches chimiquement, comme le montre le tableau 1.2, page ci-contre, mais les thiols et les sélénols diffèrent tout de même sur trois points importants : l'acidité, la polarisabilité et l'oxydabilité.

La cystéine et la sélénocystéine ont ainsi respectivement des pKa de 5,24 et de 8,53.<sup>[I.186]</sup> La conséquence de ces valeurs est qu'à pH physiologique la sélénocystéine est déprotonée. Cette propriété s'ajoute à une polarisabilité plus importante, ce qui fait de la sélénocystéine un meilleur nucléophile que la cystéine, protonée.

Propriété	Se	S
Configuration électronique	[Ar] 3d <sup>10</sup> .4s <sup>2</sup> .4p <sup>4</sup>	[Ne] 3s <sup>2</sup> .3p <sup>4</sup>
Rayon atomique (empirique)	115 pm	100 pm
Rayon de Van der Waals	190 pm	180 pm
Longueur de la liaison X-X	232 pm (Se-Se)	205 pm (S-S)
Energie de liaison X-X	332,6 kJ·mol <sup>-1</sup> (Se-Se)	425,3 kJ·mol <sup>-1</sup> (S-S)
	371,1 kJ·mol <sup>-1</sup> (Se-S)	
Longueur de la liaison C-X	196 pm (C-Se)	180 pm (C-S)
Energie de liaison C-X	590,4 kJ·mol <sup>-1</sup> (C-Se)	714,1 kJ·mol <sup>-1</sup> (C-S)
Electronégativité (Pauling)	2,55	2,58
Polarisabilité	3,8 Å <sup>3</sup>	2,9 Å <sup>3</sup>
Potentiel rédox des réactions :		
X + 2e <sup>-</sup> → X <sup>2-</sup>	-924 mV	-476 mV
RX–XR + 2e <sup>-</sup> → 2RX <sup>-</sup>	-488 mV <sup>[I.187]</sup> ou -380 mV <sup>[I.186]</sup>	-233 mV

**Tableau 1.2** – Comparaison des propriétés du sélénium et du soufre.<sup>[I.188]</sup>  
RX représente pour chaque élément l'acide aminé correspondant (sélénocystéine ou cystéine).

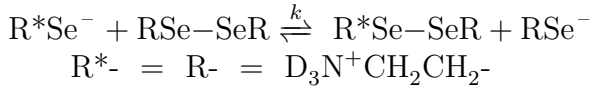
Cette entité chargée négativement et riche en électrons est sans aucun doute une cible de réaction de choix pour les molécules oxydantes. En effet, en plus d'être un bon nucléophile, le potentiel d'oxydoréduction de -380mV du couple sélénocystine/sélénocystéine<sup>[I.186]</sup> (ou de -488 mV<sup>[I.187]</sup>, selon la technique de mesure) et l'accessibilité énergétique de ses états d'oxydation intermédiaires font de la sélénocystéine un réducteur très réactif (voir Figure 1.2). En comparaison, le potentiel du couple cystine/cystéine est de -233mV.

Cela explique les données expérimentales<sup>[I.186]</sup> suivantes : une majorité des sélénocystéines catalyse des réactions d'oxydoréduction, or il a été montré que lorsque ces réactions mettent en jeu une sélénocystéine, elles sont plus rapides que celles utilisant une cystéine. En plus, la sélénocystéine permet des transferts mono- et bi-électroniques, alors que la cystéine fonctionne préférentiellement sur un mode bi-électronique.

Voici quelques exemples illustrant la plus grande réactivité de la sélénocystéine par rapport à la cystéine :

- Il est connu que les tyrosines présentes dans les protéines forment un radical tyrosyl sous l'effet d'irradiation. Il s'agit d'un radical stable, qui peut être sujet à réparation. Or la réduction de ce radical par le glutathion, assimilable à une cystéine, est lente ( $k \approx 2 \times 10^3 \text{ M}^{-1} \cdot \text{s}^{-1}$ ), comparativement à la réduction par une sélénocystéine de 5 ordres de magnitude plus rapide.<sup>[I.186]</sup>

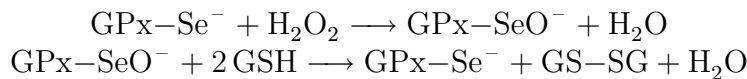
- Une étude par RMN a permis de mesurer la vitesse d'échange à laquelle se déroule la réaction d'équilibre dynamique en solution entre sélénol et diséléniure d'une amine analogue à la sélénocystéine :



La constante de vitesse est  $k = 1,65 \times 10^7 \text{ M}^{-1} \cdot \text{s}^{-1}$  pour la réaction concernant l'amine séléniée, alors que l'amine analogue soufrée s'inter-échange avec le disulfure avec une constante  $k = 68,0 \text{ M}^{-1} \cdot \text{s}^{-1}$ .<sup>[1.188]</sup> Comme les réactions d'inter-échange, ou d'échange de degrés d'oxydation, entre les différents sites oxydoréducteurs d'une enzyme telle la thiorédoxine réductase (TrxR) sont essentiels pour son fonctionnement, l'avantage que peut procurer une sélénocystéine sur la vitesse de réaction est important.

Une telle différence de vitesse de réaction sépare également les acylations et alkylations d'analogues soufrés, de celles d'analogues séléniés.

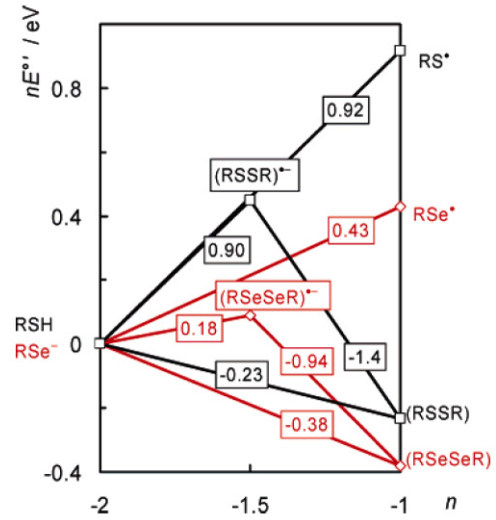
- L'exemple de la réduction du peroxyde d'hydrogène  $\text{H}_2\text{O}_2$  par la glutathion peroxydase (GPx) illustre également le rôle primordial du résidu sélénié.<sup>[1.189]</sup> Un des fonctionnements postulés, le plus répandu, de cette enzyme est le suivant : (GS représente le glutathion)



La première étape de la réaction serait une réduction du peroxyde d'hydrogène en eau par la sélénocystéine de l'enzyme, suivie de la régénération de ce site catalytique grâce au glutathion. Or une cystéine est également capable de réduire  $\text{H}_2\text{O}_2$ , mais à une vitesse de  $16\text{-}21 \text{ mol}^{-1} \cdot \text{dm}^3 \cdot \text{s}^{-1}$  contre  $5,0 \times 10^7 \text{ mol}^{-1} \cdot \text{dm}^3 \cdot \text{s}^{-1}$  pour la GPx.

Bien sûr, il faut prendre en compte le fait que l'environnement moléculaire de la sélénocystéine, au sein de la protéine, influence grandement le déroulement et la vitesse des réactions d'oxydo-réduction.<sup>[1.189]</sup> (Plusieurs ordres de magnitude séparent les vitesses de réduction de  $\text{H}_2\text{O}_2$  par la GPx et par de la sélénocystéine libre.)

Une conclusion s'impose : la sélénocystéine peut être considérée comme étant une cystéine super-réactive. Cette propriété peut également être soulignée par l'analyse des réactions que cet acide aminé catalyse, et que nous allons voir dans le paragraphe suivant.



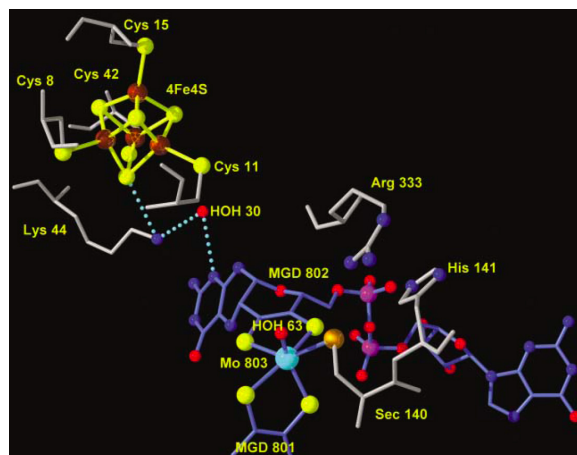
**Figure 1.24** – Diagramme d'état redox de la cystéine et de la sélénocystéine.<sup>[1.186]</sup> Les énergies sont en eV (1 eV = 96,5 kJ/mol). La pente d'une ligne de jonction entre deux composés représente le potentiel d'oxydo-réduction du couple. (ex :  $E^\circ(\text{Sec-Sec}/\text{Sec}^-) = -0,38 \text{ V}$  à pH 7 par rapport à l'électrode normale à hydrogène ENH.)

## 1.5.2 Réactions enzymatiques impliquant une sélénocystéine

### Comparaison de la réactivité entre séléno-enzymes et thio-enzymes

Les premières expériences de comparaisons d'enzymes homologues qui possèdent soit une cystéine, soit une sélénocystéine comme résidu catalytique, datent de la découverte de l'incorporation cotraductionnelle de la sélénocystéine à un codon UGA, en 1987.<sup>[1.2]</sup> Déjà, l'activité moindre d'un mutant de la formate-déshydrogénase H (FDH<sub>H</sub>) d'*Escherichia coli*, contenant une cystéine à la place de la sélénocystéine catalytique, avait été remarquée.

Cette observation fut confirmée 4 ans plus tard,<sup>[1.191]</sup> par des mesures précises des différences entre les paramètres cinétiques des deux formes de l'enzyme, qui ne diffèrent que d'un atome. Ces valeurs sont présentes dans le tableau 1.3.



**Figure 1.25** – Site actif de la FDH<sub>H</sub> d'*E. coli*.<sup>[1.190]</sup> 4Fe4S : Cluster Fer-Soufre responsable du transfert électronique. MGD : Cofacteurs molybdoptérine-guanine-dinucléotide. Mo en cyan, Se en orange. Le formate prend la place de la molécule H<sub>2</sub>O 63. (n° PDB : 1FDO)

Formes de formate-déshydrogénase	$k_{\text{cat}}$	$K_m$ (formate)	$K_d$ (formate)
[Se]FDH <sub>H</sub>	2800 sec <sup>-1</sup>	26 mM	100 mM
[S]FDH <sub>H</sub>	9 sec <sup>-1</sup>	9 mM	5 mM

**Tableau 1.3** – Comparaison des données cinétiques de la formate-déshydrogénase d'*E. coli*, dans sa version naturelle et mutée avec une cystéine à la place de la sélénocystéine.<sup>[1.191]</sup>

Ainsi, le mutant interagit plus fortement avec le substrat, comme indiqué par les plus faibles valeurs de  $K_m$  et de  $K_d$  vis-à-vis du formate, mais le débit de réaction est 300 fois moindre que pour la forme naturelle sélénée. Au final, la différence entre les rapports  $k_{\text{cat}}/K_m$  montre que l'enzyme sélénée est bien plus efficace que son analogue soufrée.<sup>[1.191]</sup> Pourtant la sélénocystéine n'est pas directement le résidu catalytique, mais influence simplement l'électrophilie du molybdène catalytique<sup>[1.192]</sup> (voir figure 1.25).

Un autre cas de réaction montrant une grande différence entre l'efficacité des enzymes sélénées et soufrées est la réaction de réduction des sulfoxydes générés par oxydation des méthionines du matériel protéique. Les enzymes s'attellant à cette tâche sont les méthionine sulfoxyde réductases MsrA et MsrB, qui réduisent stéréospécifiquement les sulfoxydes S- ou R-, respectivement.<sup>[1.193]</sup> En 2006, une étude<sup>[1.194]</sup> répertoria les sélénoprotéines MsrA identifiables dans les trois domaines du vivant. Cette étude permis de mesurer l'activité d'une MsrA de *Chlamydomona* ayant une activité de 10 à 50 fois supérieure à celle de son homologue murin, contenant une cystéine. La transformation par mutation de cette enzyme murine en sélénoprotéine fonctionnelle permit d'améliorer son efficacité, ce qui



corrobore l'intérêt catalytique de la sélénocystéine. Il est intéressant de noter que la sélénoprotéine obtenue pose tout de même le problème de ne pas pouvoir être facilement régénérée par réduction, ce que explique peut-être pourquoi seulement 14 organismes ont opté pour la sélénoprotéine MsrA.<sup>[1.194]</sup>

Des données de comparaison sont également disponibles entre la seule sélénoprotéine MsrB1 connue (ou SelR) et ses homologues soufrés<sup>[1.196,1.197]</sup> : ainsi cette enzyme mammifère est 100 fois plus active que sa version cystéinylée mutante. Cette observation fonctionne également en sens inverse : lorsque les MsrB2 et MsrB3 furent mutées pour y introduire une sélénocystéine à la place de leur cystéine catalytique, l'activité augmenta d'un facteur 100.

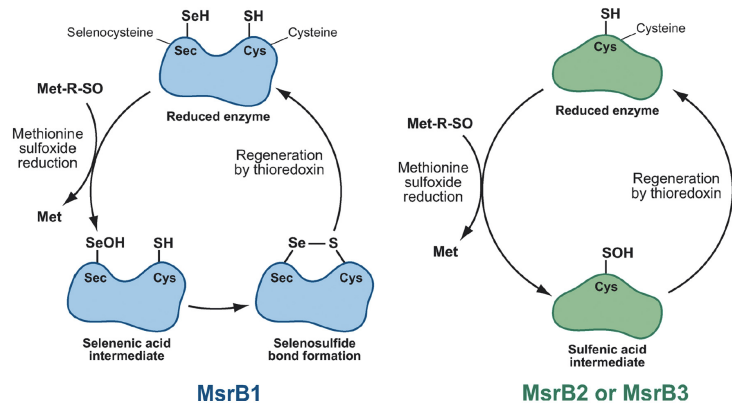


Figure 1.26 – Principe de fonctionnement des MsrB.<sup>[1.195]</sup>

Un troisième système a été fortement étudié pour découvrir la différence d'efficacité entre protéines utilisant la cystéine et la sélénoprotéine : les thiorédoxine réductases (TrxR).<sup>[1.187,1.198-1.202]</sup> Les thiorédoxine réductases mammifères sont des dimères, dont chaque monomère possède un corps portant un couple de cystéines oxydoréductrices et une séquence C-terminale conservée -Gly-Cys-Sec-Gly. Cette pénultième sélénocystéine est essentielle à l'activité rédox de l'enzyme : chez le rat, le remplacement de cette sélénocystéine par une cystéine provoque une diminution du  $k_{cat}$  d'un facteur 100 et le  $K_m$  d'un facteur 10, mais l'enzyme reste active.<sup>[1.199]</sup> Il en va de même avec un mutant de la TrxR humaine placentaire.<sup>[1.198,1.201]</sup> Un mécanisme réactionnel, impliquant un trio Sec-His-Glu, a été proposé.<sup>[1.203]</sup>

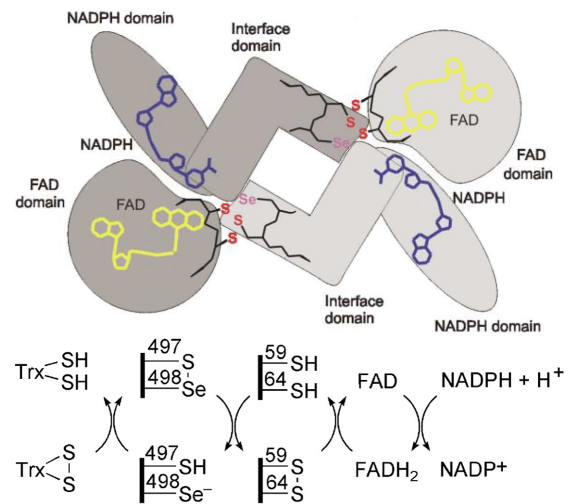


Figure 1.27 – Schéma de l'organisation du dimère de TrxR, et réaction de réduction de la thiorédoxine aux dépens du NADPH.<sup>[1.9,1.198]</sup>

Le quatrième exemple de différence d'activité entre une cystéine et une sélénocystéine, concerne une enzyme ayant une activité très différente des précédentes : les déiodinases de type I, II et III (D1, D2 et D3) catalysent une réaction impliquant T4 (3,5,3',5'-tetraiodothyronine), le précurseur de l'hormone thyroïdienne T3.<sup>[1.204]</sup> Ainsi, D1, officiant dans le foie et les reins, et D2, officiant dans le cerveau, catalysent la déiodination de T4 en T3 (3,5,3'-triiodothyronine), régulant ainsi la concentration de cette hormone, respectivement dans le plasma et dans le cerveau. À l'inverse, D3, enzyme dont l'activité est

foétale et placentaire, catalyse la déiodination de T4 en rT3 (3,3',5'-triiodothyronine) métabolite inactif, et de T3 en T2 (3,3'-diiodothyronine) également inactif, dans un possible but de protéger les organes foétaux d'une exposition à de trop fortes concentrations de cette hormone. Ces deux réactions d'inactivation sont aussi catalysées par D1, ainsi que la déiodination de rT3 en T2.

La mutation (Sec→Cys) provoque des résultats similaires chez toutes ces sélénoprotéines : une augmentation du  $K_m$  vis-à-vis des substrats et une diminution de l'activité spécifique correspondante, selon les proportions présentées dans le tableau 1.4.

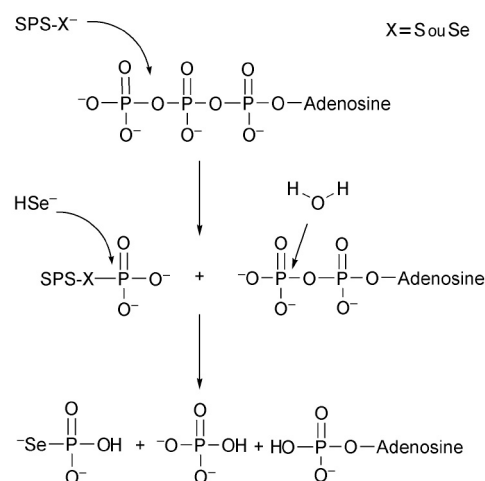
Type de déiodinase	substrat	Activité spécifique	$K_m$
D1 <sub>(Cys)</sub> par rapport à D1 <sub>(Sec)</sub>	rT3 <sup>[I.205]</sup>	divisée par 100	multipliée par 3
D2 <sub>(Cys)</sub> par rapport à D2 <sub>(Sec)</sub>	T4 <sup>[I.206]</sup>	divisée par 10	multipliée par 1000
D3 <sub>(Cys)</sub> par rapport à D3 <sub>(Sec)</sub>	T4 <sup>[I.204]</sup> T3 <sup>[I.204]</sup>	divisée par 6 divisée par 2	multipliée par 100 multipliée par 5

**Tableau 1.4** – Conséquence d'une mutation (Sec→Cys) chez les déiodinases.<sup>[I.204]</sup>

En parallèle à cette perte d'activité, la protéine mutée est exprimée de 20 à 400 fois plus que la protéine native,<sup>[I.205]</sup> selon la lignée cellulaire choisie pour la production. Cela peut être expliqué en partie par la proportion de terminaisons résiduelles de la traduction au codon UGA lors de la traduction des gènes des sélénoprotéines. Mais on peut alors se demander si l'utilisation par les organismes de ces sélénoprotéines procure réellement un avantage, par leur efficacité supérieure, malgré une expression moindre.

Le dernier exemple illustre un cas unique où il n'y a pas d'avantage probant à l'utilisation d'une sélénoprotéine par rapport à son homologue soufré : celui des sélénophosphate synthétases. Ces enzymes, qui catalysent la formation de sélénophosphate à partir de séléniure et d'ATP, ne présentent pas de différence de comportement entre leur version séléniée et soufrée.<sup>[I.207, I.208]</sup> En effet, G. M. Lacourcière *et al* montrèrent qu'il n'y avait pas de différence flagrante de réactivité entre la sélénophosphate synthétase SelD d'*Escherichia coli* et la SelD homologue séléniée de *Haemophilus influenzae*,<sup>[I.208]</sup> comme indiqué dans le tableau 1.5, page suivante.

Cette réaction de synthèse de sélénophosphate à partir de séléniure et d'ATP n'est pas catalysée grâce au caractère oxydoréducteur du sélénium ou du soufre, mais grâce à d'autres propriétés du site actif. Il est considéré qu'un thiol ou un séléniol activés est le minimum requis pour cette réaction.<sup>[I.9]</sup> Cela pourrait expliquer que, dans ce cas, les deux composés se comportent comme les homologues chimiques qu'ils sont.



**Figure 1.28** – Mécanisme supposé du fonctionnement de la sélénophosphate synthétase (SPS).<sup>[I.9]</sup>

Une connaissance plus précise du mécanisme de réaction de SelD serait d'un grand intérêt pour l'éclaircissement de ce phénomène.

Séléno-phosphate synthétase	Activité spécifique	$K_m$ (ATP)	$K_m$ (séléniure)
SelD d' <i>E. coli</i>	29 nmol/min par mg d'enzyme	0,9 mM	20 mM
SelD de <i>H. influenzae</i>	16 nmol/min par mg d'enzyme	1,3 mM	25 mM

**Tableau 1.5** – Comparaison des données cinétiques de la séléno-phosphate synthétase d'*E. coli*, contenant une cystéine en position 16, et de *H. influenzae*, dont le 16<sup>ème</sup> résidu est une sélénocystéine.<sup>[1.208]</sup> Les données mesurées par Z. Veres en 1994<sup>[1.209]</sup> donnaient une activité spécifique de 83 nmol/min par mg d'enzyme ( $k_{cat} = 3 \text{ min}^{-1}$ ) pour SelD d'*E. coli*.

Les grandes différences observées, la plupart du temps, entre les sélén-enzymes et leurs homologues cystéinylées, permettent de comprendre pourquoi certains organismes préfèrent conserver la voie complexe de synthèse et d'incorporation de cet acide aminé au site actif de protéines. Pourtant, dans tous ces cas, la sélénocystéine peut être substituée par une cystéine, biologiquement plus accessible, bien que cela entraîne une perte d'activité. La sélénocystéine confirme son statut de cystéine "super-réactive".

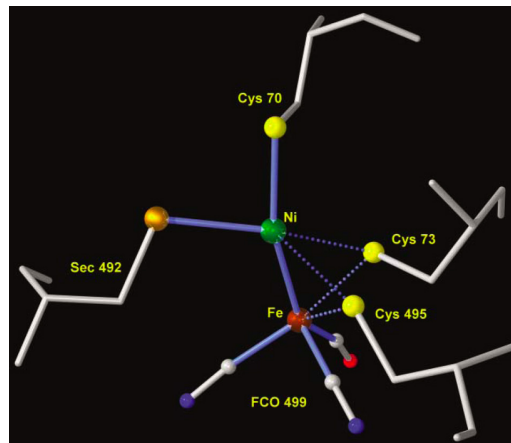
### Autres réactions impliquant une sélénocystéine

Il existe d'autres protéines où la sélénocystéine (ou la cystéine) constitue un ligand pour la coordination d'ions métalliques.

Ainsi les thiols présents dans certaines protéines peuvent être des ligands de coordination d'une large gamme d'ions métalliques comme le fer, zinc, cadmium, mercure, cobalt et cuivre, mais aucun ion du groupe 1 ou 2 (alcalin et alcalino-terreux). La coordination thiol-zinc est même utilisée de façon importante dans les motifs structuraux à "doigt de zinc".

À l'inverse, les sélénocystéines ne sont retrouvées que rarement dans une coordination métallique. Quelques exemples sont néanmoins célèbres, comme la formate déshydrogénase H d'*E. coli* dont nous avons vu le principe précédemment. Il arrive donc que la sélénocystéine soit coordonnée à un ion molybdène, ou à un ion tungstène, comme dans l'exemple de la formate déshydrogénase de *Desulfovibrio gigas*.<sup>[1.187]</sup>

Une sélénocystéine peut aussi se retrouver dans des hydrogénases particulières contenant du fer et du nickel, se retrouvant chez des organismes de la famille *Desulfovibrio*.<sup>[1.211]</sup> La réaction qu'elles catalysent est la réduction réversible de deux protons en dihydrogène. Dans ces [NiFeSe] hydrogénases, le nickel est coordonné à la sélénocystéine et à trois cystéines et forme une liaison avec le fer, qui lui complexe CN et CO dans la structure

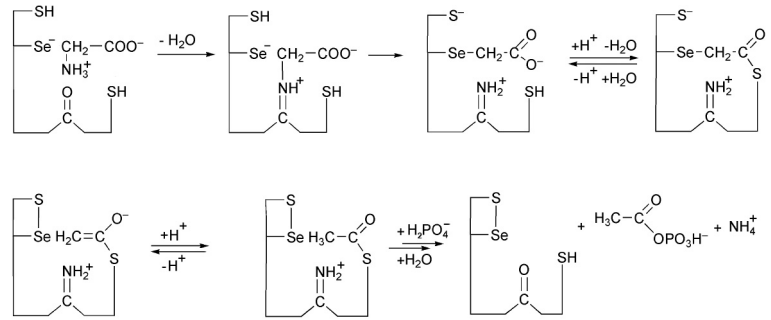


**Figure 1.29** – Site actif de la [NiFeSe] hydrogénase de *Desulfo-microbium baculatum*.<sup>[1.210]</sup> Se en orange, Ni en vert, Fe en brun sous forme complexé à deux cyanure et un monoxyde de carbone FCO 499. (n° PDB : 1CC1)

cristallographique<sup>[I.210]</sup> (voir figure 1.29, page précédente). Ces hydrogénases sont des assemblages complexes hétéro-oligomériques. Malgré la connaissance structurale du site actif, le rôle du sélénium dans la réaction n'est pas connu, la réaction se déroulant également convenablement grâce à des [NiFe] hydrogénases, sans sélénium.

Enfin, une sélénoprotéine très particulière a également été décrite : la glycine réductase est un cas de protéine où la sélénocystéine ne peut pas être remplacée par un autre acide aminé, ni même une cystéine, dans le rôle qu'elle occupe dans le site actif.

Dans la réaction de formation d'ATP à partir de glycine, la glycine réductase catalyse la première étape, produisant comme intermédiaire de l'acétyl-phosphate, qui permet d'obtenir de l'ATP par l'action de l'acétate kinase. Le processus de réaction a été décrit pour la protéine A du complexe glycine réductase de *Clostridium*<sup>[I.212]</sup> (voir le schéma 1.30).



**Figure 1.30** – Mécanisme proposé lors de la réaction catalysée par la glycine réductase de *Clostridium*.<sup>[I.212]</sup>

Nous venons de voir en quoi la sélénocystéine est un acide aminé essentiel pour l'efficacité de certaines réactions, vitales pour la cellule. Pourtant, de nombreux organismes ont réussi à se passer de cet acide aminé, très demandant en ressource, comme d'autres ont acquis, par transfert horizontal de gène, la possibilité de synthétiser des sélénoprotéines leur conférant un avantage. Nous allons voir dans la section suivante ce que la dynamique de ces pertes/gains de capacité à synthétiser les sélénoprotéines peut nous apprendre sur l'histoire et l'évolution de leur utilisation.

## 1.6 Implications évolutives

Le nombre de sélénoprotéines codées dans le matériel génétique varie selon les organismes : l'homme en possède 25, la souris 24, les organismes aquatiques une trentaine et, parmi les eubactéries, certains organismes possèdent jusqu'à 31 sélénoprotéines (par ex. *Syntrophobacter fumaroxidans*<sup>[I.165]</sup>). Un tel panel de sélénoprotéines permet d'amortir l'utilisation du système de production complexe de la sélénocystéine dans les cellules.

De façon surprenante, des animaux ne possédant qu'un seul gène de sélénoprotéine dans leur génome ont été identifiés.<sup>[I.213]</sup> Ainsi, *Caenorhabditis elegans* et *C. briggsae* ne produisent que la sélénoprotéine thiorédoxine réductase, pour laquelle ils conservent la complexe machinerie de synthèse de la sélénocystéine. D'autres nématodes possèdent des sélénoprotéines supplémentaires, retrouvées chez les autres eucaryotes. Ces données

suggèrent une réduction du sélénoprotéome durant l'évolution de ces espèces, avec comme résultat la subsistance d'un unique codon UGA recodé en sélénocystéine.

L'élément SECIS de cette unique recodage étant le seul à interagir avec la machinerie d'insertion de la sélénocystéine, il a évolué vers une forme légèrement différente de celle que l'on trouve chez les autres eucaryotes, avec une guanine à la place d'une adénine conservée.<sup>[1.213]</sup> Cela indique une capacité d'évolution du système d'insertion de la sélénocystéine, lorsque celui-ci n'est pas sujet à la pression due à son partage dans l'expression de plusieurs sélénoprotéines, et à contrario, une certaine stabilité de ce système lorsque le panel de sélénoprotéines est important.

D'autres organismes ne possèdent aucune sélénoprotéine codée dans leur génome : 80% des eubactéries séquencées, les végétaux supérieurs (ou végétaux vasculaires), les champignons et même certains animaux, parmi les insectes. Cette observation conduit à l'interrogation suivante :

La comparaison du mode de synthèse et d'incorporation de la sélénocystéine parmi les trois domaines du vivant permet de déduire d'une origine commune, ancestrale, puis une évolution divergente de ces mécanismes. Dans ce cas, comment expliquer la très grande disparité d'utilisation de la sélénocystéine parmi les organismes actuels, et la perte de cette capacité chez certains d'entre eux ?

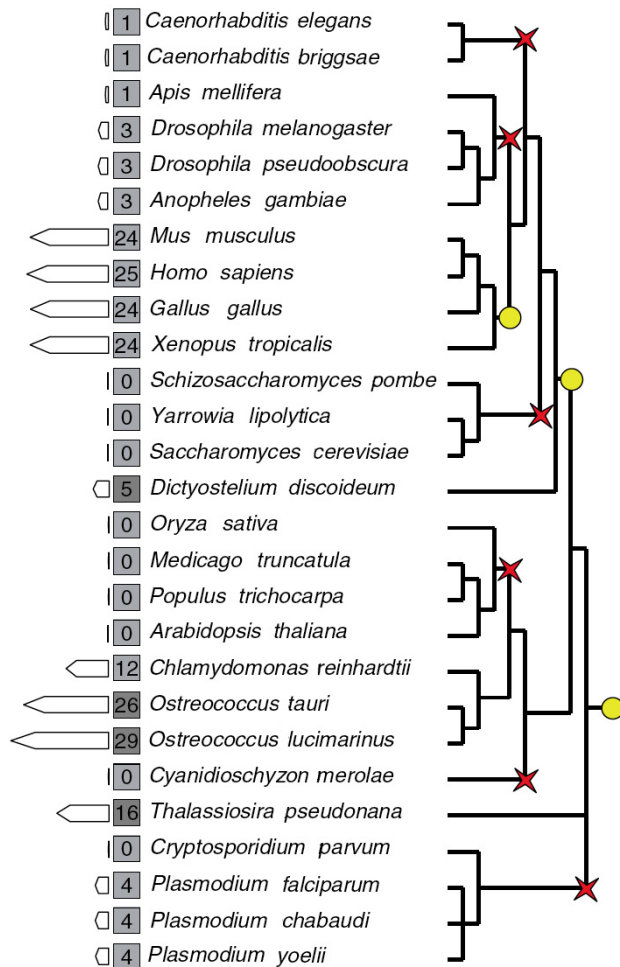
La compréhension de ce phénomène peut être obtenue par l'étude du parallèle entre cette disparité d'utilisation des sélénoprotéines, et les différences dans l'environnement de vie (ressources en sélénium, mode de vie) et l'embranchement phylogénétique auquel appartient l'organisme.

### 1.6.1 Les sélénoprotéines Archaeas

Peu de données sont disponibles sur les sélénoprotéines archaeas.<sup>[1.130]</sup> Dans ce domaine du vivant, les seuls organismes connus utilisant la sélénocystéine sont des méthanogènes qui dépendent entièrement de la disponibilité de sélénium dans leur milieu (*Methanococcus jannaschii*), ou croissent mieux lorsqu'il est présent (*Methanococcus voltae*, *M. vannieli*, *Methanopyrus kandleri*...)<sup>[1.9]</sup> *M. jannaschii* possède 7 sélénoprotéines : une sélénophosphate synthétase et 6 séléno-enzymes impliquées dans le cycle méthanogène de réactions de réduction du CO<sub>2</sub> en méthane, qui fait partie du métabolisme énergétique de cette espèce. Les autres méthanogènes cités précédemment possèdent des sélénocystéines homologues, impliquées dans les mêmes réactions, les sélénoprotéines absentes de leur génome étant remplacées par des homologues cystéinylés.

Le manque de données concernant les archaeas constitue un biais, qui résulte en une occurrence très spécifique de l'utilisation de sélénoprotéines dans ce domaine du vivant. Néanmoins, cela peut faire naître un doute quant à l'universalité originelle de l'utilisation de la sélénocystéine, et va dans le sens d'une émergence ou d'une acquisition localisée du recodage de cet acide aminé dans ce domaine.

## 1.6.2 Une grande disparité des sélénoprotéomes eucaryotes



**Figure 1.31** – Cladogramme simplifié d'eucaryotes modèles. Les flèches de gauche sont proportionnelles au nombre de sélénoprotéines qui sont codées dans leurs génomes (nombre encadré).<sup>[1.77]</sup>

dans des groupes d'organismes (symbolisées par des croix rouge sur le cladogramme) et à la possible émergence de nouvelles familles dans d'autres branches (symbolisée par des cercles jaunes sur le cladogramme).

Ainsi, il est remarquable que les grands sélénoprotéomes sont réservés aux organismes de grandes tailles et à certains organismes aquatiques. Les études de dynamique évolutive des sélénoprotéomes parmi les eucaryotes<sup>[1.77]</sup> sont formelles sur l'émergence ancestrale de la plupart de ces protéines, ainsi que sur les massives pertes de sélénoprotéines dans des branches eucaryotiques indépendantes, quasiment toutes terrestres.

Il est donc supposé que la vie aquatique est en faveur de l'utilisation massive de sélénoprotéines, alors que l'habitat terrestre conduit, à l'exception des vertébrés, à une réduction de la dépendance à ces protéines pour le fonctionnement cellulaire.

La distribution des gènes de sélénoprotéines dans les génomes eucaryotes est très disparate. Comme précisé précédemment, leur nombre varie de plus de trente à zéro.

Le cladogramme 1.31 illustre bien cette disparité.<sup>[1.77]</sup>

Les eucaryotes ne possédant pas de gène de sélénoprotéine font partie des fungi (ou mycota), des végétaux supérieurs (ou embryophyta), des algues rouges (ou rhodophyta) et de certaines familles de protistes.

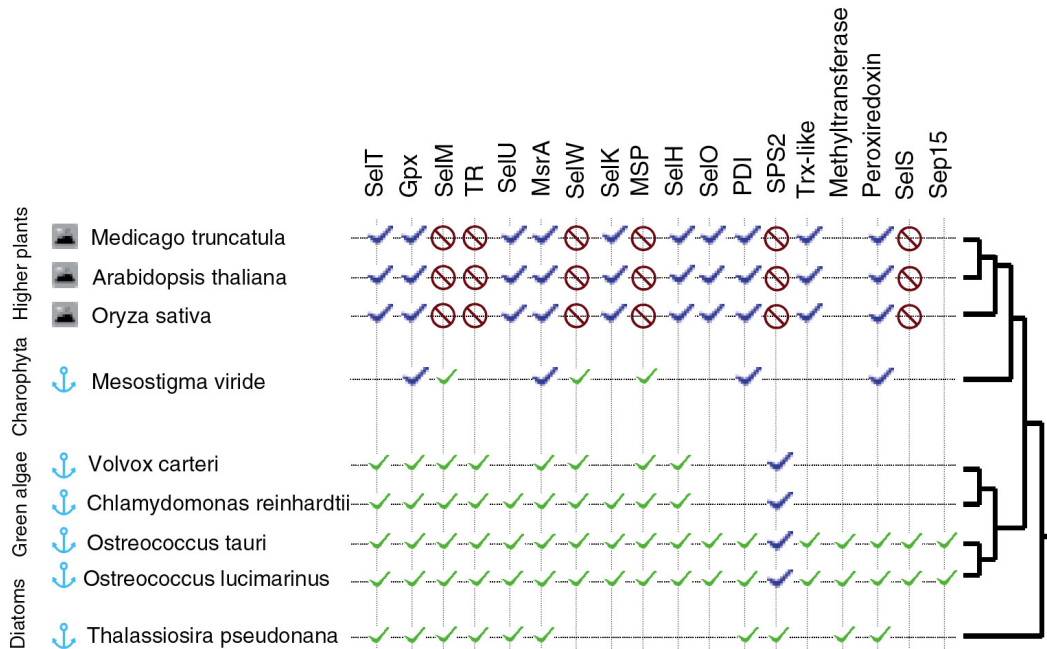
Ceux possédant peu de gènes de sélénoprotéines (moins de 10) font partie des nématodes, des insectes, des amibes sociales (ou myxomycètes) et également de familles de protistes.

Enfin, les eucaryotes possédant un nombre important de sélénoprotéines sont les vertébrés, les algues vertes (ou chlorophyta) et les diatomées.

Toutes ces sélénoprotéines eucaryotiques appartiennent à un nombre restreint de familles. La distribution de ces familles au sein de la phylogénie eucaryotique permet de conclure à la

perte de certaines familles de protéines

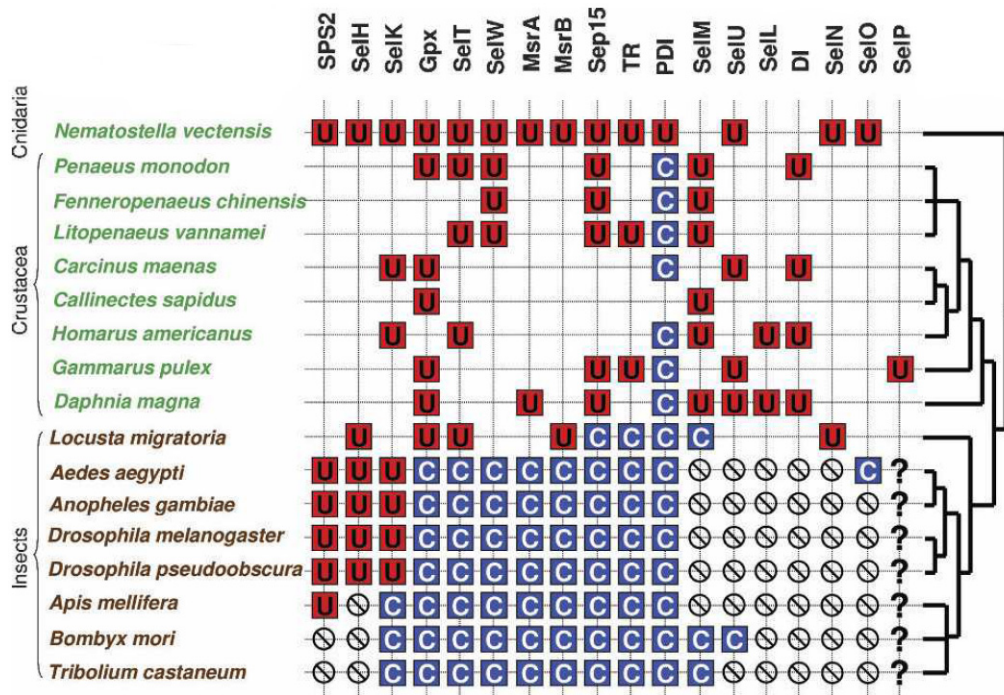
Ces hypothèses sont corroborées dans les exemples que constituent le règne des plantes et la branche des arthropodes, illustrés par les schémas 1.32 et 1.33.



**Figure 1.32** – Occurrence des sélénoprotéines chez les plantes.<sup>[1.77]</sup> La présence de sélénoprotéines est symbolisée en vert, celle de leurs homologues cystéinylées en bleu. Certains génomes ne sont pas entièrement séquencés (cases vides).

Ainsi, les plantes terrestres utilisent des protéines où une cystéine occupe la place de la sélénocystéine active. Il est à noter que les végétaux supérieurs ont également perdu la machinerie nécessaire à l’incorporation de la sélénocystéine. Au cours de l’évolution, le passage d’une vie aquatique ancestrale à une vie terrestre a marqué les plantes ayant effectué ce changement de milieu, par une substitution de leur sélénoprotéome en “cystéinoprotéome”, par rapport à celles dont l’habitat est resté aquatique.

De façon parallèle, la branche des arthropodes a également été marquée, mais de façon moins contrastée, dans le passage à une vie terrestre.<sup>[1.101]</sup> (voir figure 1.33, page ci-contre) À l’exception de deux insectes, *Tribolium castaneum* et *Bombyx mori*, les arthropodes terrestres ont conservé l’usage des sélénoprotéines, mais en nombre restreint. La plupart des sélénoprotéines ont là aussi été remplacées par des homologues cystéinylées, certaines familles de sélénoprotéines ayant même été simplement perdues, comme SelH, SelU et SelM. Les insectes sans sélénoprotéines ont également perdu en partie la machinerie de synthèse et d’incorporation de la sélénocystéine.



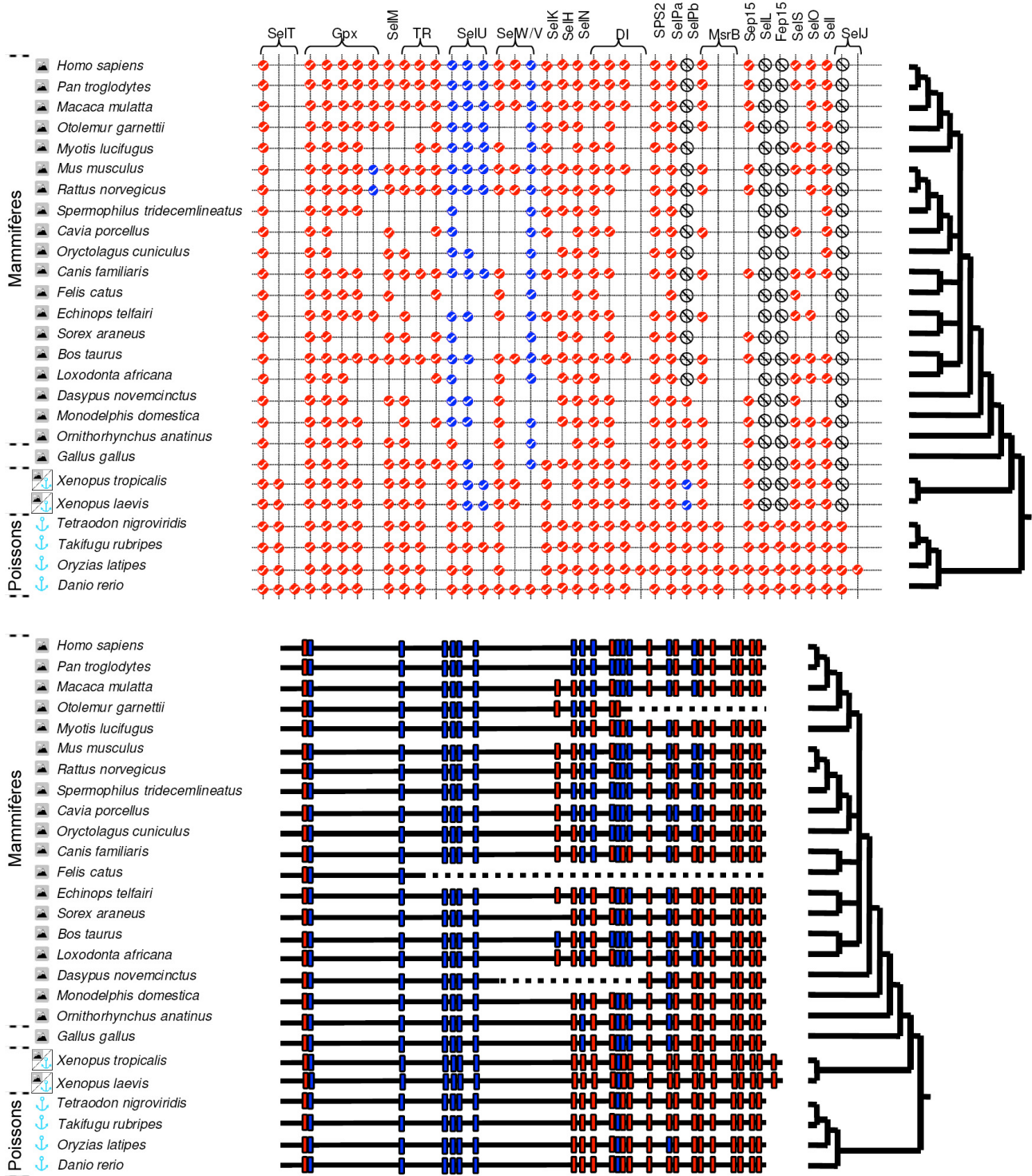
**Figure 1.33** – Occurrence des sélénoprotéines chez les arthropodes, par rapport à un organisme de référence.<sup>[1.101]</sup> Les sélénoprotéines sont symbolisées par des U, leurs homologues cystéinyliées par un C. Certains génomes ne sont pas entièrement séquencés (cases vides).

La branche des eucaryotes dont les sélénoprotéomes ont été les mieux conservés malgré une transition terrestre, sont les vertébrés et en particulier les mammifères. La raison pour laquelle ils n’ont pas été affectés au même titre que les plantes et les arthropodes est aujourd’hui un mystère.

Certaines observations intéressantes ont néanmoins pu être effectuées sur les sélénoprotéomes mammifères par comparaison avec ceux, plus importants encore, d’eucaryotes marins (jusqu’à 34 sélénoprotéines pour *Danio rerio*, le poisson zèbre). En plus d’une tendance générale “aquatique/terrestre” de 32-34 sélénoprotéines pour les poissons contre 23-25 pour les mammifères, il faut remarquer que certaines sélénoprotéines ont été remplacées par leurs homologues cystéinyliées chez tous les mammifères, comme SelU, et que d’autres familles ont simplement disparu, comme SelJ, SelL, Fep15 et SelPb (voir le haut de la figure 1.34, page suivante).

Un parallèle surprenant a pu être dévoilé lors de l’étude de ces sélénoprotéomes. En effet, il semble qu’il y ait une corrélation entre la taille du sélénoprotéome d’une espèce et le contenu en sélénocystéine de leur sélénoprotéine P (SelP).<sup>[1.214]</sup>





**Figure 1.34** – Comparaison, chez 19 mammifères, 1 oiseau, 2 amphibiens et 4 poissons : [en haut] de l'occurrence de sélénoprotéine (en rouge) et de leurs homologues cystéinyliées (en bleu). Les génomes des organismes présentés ont été entièrement séquencés. [en bas] de la répartition et de la position des sélénocystéines (en rouge) et des cystéines (en bleu) dans un alignement de leurs protéines SelPa.<sup>[1.214]</sup>

Nous avons vu rapidement à la section 3, page 29, que SelP est une protéine que l'on retrouve dans la circulation, directement liée à l'approvisionnement en sélénium de certains organes. Il n'est alors pas étonnant que les poissons, utilisant plus de sélénoprotéines, aient besoin d'un approvisionnement plus important de leurs organes en sélénium, et qu'ils utilisent deux versions de la protéine SelP (SelPa et SelPb). De plus, la comparaison des gènes de SelPa entre les mammifères et les poissons met en relief un contenu beaucoup plus riche en sélénocystéine des SelP de poissons, et un fort remplacement unidirectionnel de ces sélénocystéines par des cystéines dans les SelP mammifères (voir le bas de la figure 1.34, page ci-contre). Une expression plus importante de cette protéine et un taux supérieur de sélénium dans les tissus ont même été mesurés chez les poissons.

Les différences entre les tailles des sélénoprotéomes dans le domaine eucaryote suivent donc une certaine logique phylogénique, et l'importance qu'un organisme donne au sélénium est répercutée jusque dans le parcours physiologique de cet élément. Quelles sont donc les raisons de ces phénomènes ?

### 1.6.3 Hypothèses sur les facteurs entraînant cette disparité

Des explications ont été avancées pour expliquer cette grande différence de sélénoprotéomes entre les organismes aquatiques et terrestres.<sup>[I.214]</sup>

En suivant un modèle de décroissance des sélénoprotéomes à travers l'évolution, un premier facteur pouvant conduire à la perte de sélénoprotéines est une plus mauvaise disponibilité de sélénium en habitat terrestre qu'en habitat aquatique où le sélénium est constamment disponible, conduisant à une pression génétique favorisant la perte des caractères demandeurs en sélénium chez les organismes terrestres.

Un deuxième facteur possible est lié à la plus forte réactivité du sélénium, son avantage lors de son utilisation en catalyse enzymatique. En milieu terrestre, les organismes sont plus exposés à l'oxygène et au rayonnement U.V., qui provoquent tous deux facilement des dommages aux sélénoprotéines, par oxydation et formation de radicaux. Les organismes de taille plus importante, comme les mammifères, possèdent un système efficace de défense contre ces agressions, ce qui peut expliquer une perte moindre de leur sélénoprotéome.

Ces propositions sont à mettre en relation avec les hypothèses concernant l'existence du codage de la sélénocystéine par le codon UGA, et l'émergence de ce codage.<sup>[I.215]</sup>

Une première hypothèse propose le scénario suivant<sup>[I.216]</sup> : le code génétique comportait initialement la sélénocystéine codée par le codon UGA, dans les organismes primitifs anaérobies. Comme l'atmosphère devint de plus en plus riche en oxygène, grâce à l'émergence de la photosynthèse chez certains organismes, l'utilisation de la sélénocystéine devint un caractère désavantageux, ce qui provoqua rapidement sa moindre utilisation et la diminution des sélénoprotéomes, jusqu'à sa disparition dans certains organismes.

Une seconde hypothèse prévoit l'addition tardive de la sélénocystéine au code génétique, et son utilisation dans un nombre restreint de sélénoprotéines primitives.<sup>[I.217]</sup> Cette

hypothèse conduit à un scénario d'augmentation du nombre de sélénoprotéines jusqu'à son apogée, puis à un déclin de celui-ci lors du passage à une vie terrestre, avec le maintien chez les eucaryotes terrestres des sélénoprotéines ayant des fonctions antioxydantes.

Les facteurs environnementaux semblent avoir influencé l'évolution de l'utilisation de la sélénocystéine et la façon dont celle-ci est synthétisée puis incorporée. Aux facteurs déjà cités, nous pouvons ajouter la rareté relative de l'élément sélénium, qui a pu conduire originellement à la mise en place d'un système d'"antiblocage" de la machinerie cellulaire lors de l'expression des sélénoprotéines, par le biais d'une possibilité de terminaison de la traduction lors de l'insuffisance de cette ressource. Cette émergence de la double signification du codon UGA a pu être un avantage pour les organismes possédant une telle machinerie lors de la raréfaction locale de cet élément. De la même façon, cette rareté a pu conduire à l'utilisation d'enzymes alternatives. La combinaison de ces deux mécanismes peut expliquer la facilité de la perte totale de l'utilisation de cet élément sous forme de sélénocystéine chez une majorité d'organismes.

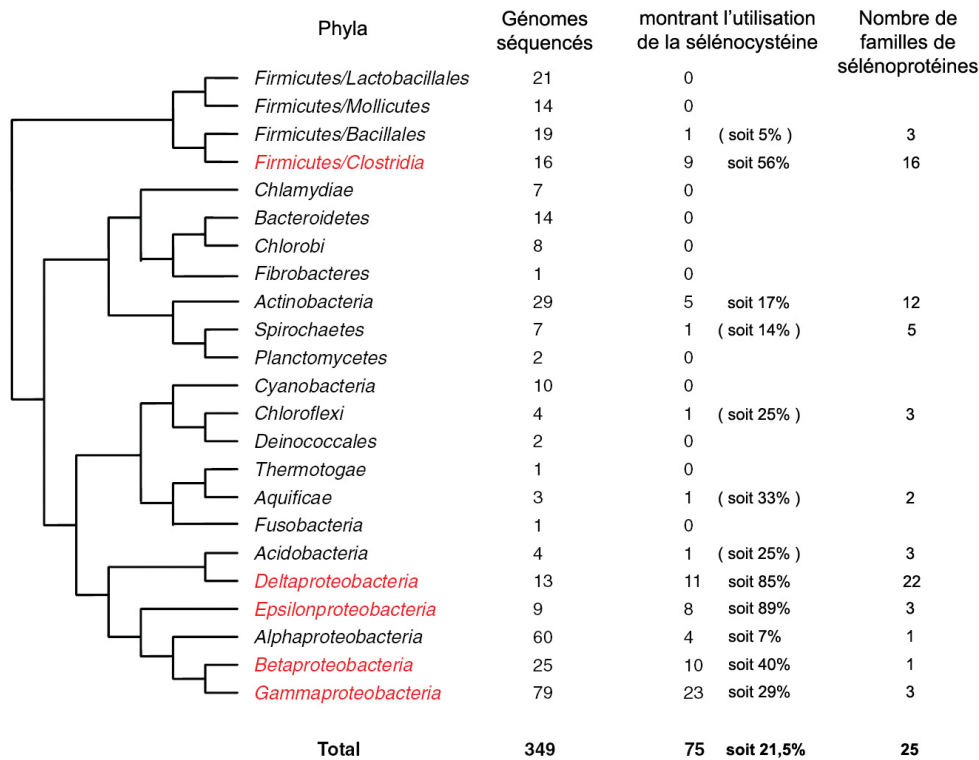
Dans tous les cas, il n'est pas surprenant que le codon "choisi" pour coder la sélénocystéine soit UGA.<sup>[1.215]</sup> Ce codon est celui qui peut prendre le plus de fonctions différentes dans le vivant : en plus de pouvoir être un codon de terminaison, il sert de codon pour la sélénocystéine, pour la cystéine chez *Euplotes octocarinatus*, pour le tryptophane chez les mitochondries, *Mycoplasma* et *Sprrioplasma* et de façon inefficace chez *Bacillus subtilis*.

Si la sélénocystéine avait été ajoutée au code génétique de façon tardive, le codon UGA, programmé de façon lâche dans l'évolution, aurait été le candidat de choix pour cet ajout. De même, si la sélénocystéine avait été codée précocement de façon plus ou moins ambiguë, il ne serait pas surprenant que le codon libéré par la perte de l'utilisation des sélénoprotéines chez certains organismes conduise à une reprogrammation lâche de ce codon.

#### 1.6.4 Confirmation de ces tendances par l'étude des eubactéries

L'étude de l'usage de la sélénocystéine chez les eubactéries et sa mise en parallèle avec les conclusions précédentes tirées de l'étude des eucaryotes peut permettre de confirmer ou infirmer ces hypothèses et peut-être enfin de découvrir le véritable scénario de l'émergence de la sélénocystéine, de la divergence des mécanismes de synthèse et d'incorporation de cet acide aminé, et des relations entre les sélénoprotéomes et leurs tendances d'évolution.

Des données sont déjà actuellement disponibles au sujet des sélénoprotéomes bactériens : seules 20% des eubactéries séquencées possèdent aujourd'hui la capacité de synthétiser la sélénocystéine et les sélénoprotéines. Leurs répartitions dans la phylogénie des procaryotes n'est ni homogène, ni très localisée.<sup>[1.91]</sup> (Voir l'arbre phylogénique présent en Annexe A.1, page 229.) Les sélénoprotéomes des eubactéries comportent de 1 à 31 sélénoprotéines, réparties dans seulement 25 familles. Il est à noter que les familles d'organismes possédant les protéomes les plus importants sont surtout les  $\delta$ -protéobactéries et les firmicutes, sous-branche clostridia.



**Figure 1.35** – Répartition phylogénique des bactéries utilisant la sélénocystéine.<sup>[I.165]</sup>

Cet “état des lieux” soulève une première contradiction entre l’avantage catalytique apporté par la sélénocystéine et le nombre faible d’organismes utilisant cet acide aminé.

Les études comparatives des sélénoprotéomes et des gènes *selABCD* bactériens ont permis à Y. Zhang, H. Romero, G. Salinas et V. N. Gladyshev d’apporter des pistes de réponses à cette interrogation.<sup>[I.91,I.165]</sup> Leurs données montrent tout d’abord le fort dynamisme de l’évolution de l’utilisation des sélénoprotéines dans le domaine bactérien, qui peut être illustré par les faits suivants :

- L’existence ancestrale de l’utilisation de la sélénocystéine permet de déduire que, si 80% des eubactéries séquencées ont perdu ce caractère aujourd’hui, une proportion plus importante d’eubactéries a subi ce phénomène de perte et une proportion d’entre elles ont regagné ce caractère par transfert de gènes horizontal. Ce fait peut être déduit de l’analyse de la phylogénie obtenue par comparaison des gènes *selABCD*, mettant en relief les transmissions verticales et horizontales de ces gènes.<sup>[I.165]</sup> Un exemple d’arbre phylogénique illustrant cette transmission est disponible, pour *SelA*, en Annexe A.3, page 232. La perte ou le gain de ce caractère ne sont donc pas irréversibles.

- Chez des organismes possédant un grand sélénoprotéome, l’émergence de nouvelles familles de sélénoprotéines par mutations Cys→Sec a pu être montrée (7 nouvelles familles possédant moins de 4 membres). En plus de ces émergences, 10 familles de sélénoprotéines ont bénéficié de 21 mutations Cys→Sec. Au total, 15 familles sur 17 sont des familles de protéines oxydoréductrices basées sur les motifs UxxC, CxxU ou TxxU.

À l'inverse, seules 15 mutations Sec→Cys dans 4 familles de sélénoprotéines ont pu être montrées, et seuls 4 de ces événements ont conduit à la perte de sélénoprotéomes. Des structures de type SECIS ont même été trouvées en aval des codons Cys correspondants. Deux tiers de ces mutations concernent des protéines SelD ou FdhA, qui constituent pourtant les familles de sélénoprotéines se retrouvant chez le plus grand nombre d'organismes.

Ce fort dynamisme évolutif de l'utilisation de la sélénocystéine pourrait rendre l'hypothèse d'une émergence localisée, puis d'une diffusion de cette capacité, attractive. Une simple corrélation entre les distances phylogéniques séparant les ARN 16S de la petite sous-unité des ribosomes des différentes espèces et celles séparant les protéines SelA, SelB et SelD de ces espèces, permet de vérifier l'origine ancestrale de cette capacité d'utilisation de la sélénocystéine (voir les 3 graphiques à gauche de la figure 1.36, page suivante).

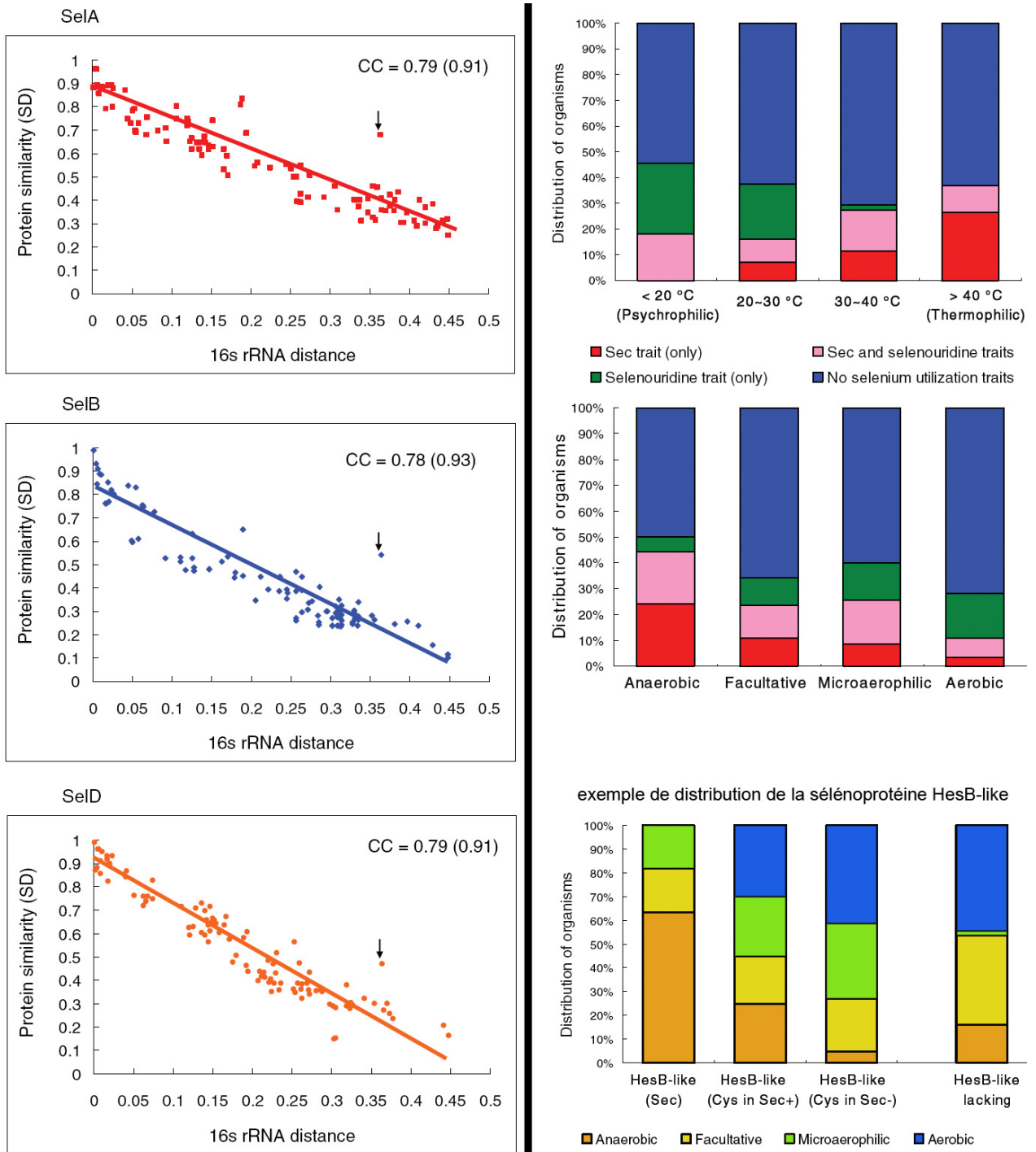
En effet, cette corrélation montre bien que l'évolution de la machinerie de synthèse et d'insertion de la sélénocystéine suit la divergence des espèces, à l'exception des rares phénomènes de transferts horizontaux de gènes entre espèces. Ce système a donc généralement bien été hérité par filiation à un ancêtre commun.

Le phénomène de transfert de gènes ne concerne pas seulement la machinerie permettant l'utilisation de la sélénocystéine, mais également des sélénoprotéines. Un transfert entre des archaées et des eubactéries a même pu être montré : *Syntrophobacter fumaroxidans*, une  $\delta$ -protéobactérie possédant le plus grand des sélénoprotéomes bactériens, possède dans son génome de multiples copies d'un opéron *glpC-hdrA-frhD-frhG-frhA*, absent chez les organismes apparentés (HdrA, FrhD et FrhA sont des sélénoprotéines). Un opéron quasi-identique est présent chez les archaées méthanogènes capable d'utiliser la sélénocystéine. Un tel transfert est envisageable, *Syntrophobacter fumaroxidans* étant un anaérobie vivant en association syntrophique avec des méthanogènes. Une interrogation concernant la modalité de ce transfert est pourtant soulevée par l'utilisation de différents mécanismes de recodage du codon UGA chez les eubactéries et archaées actuelles. Mais aucune forme de structure SECIS d'archaées n'a pu être retrouvée au niveau des opérons présents dans la bactérie, et réciproquement.

Une importante conclusion s'impose alors : ce transfert "récent", confirmé par les analyses d'identité de gènes, met en évidence la possibilité d'une compatibilité, même ponctuelle, entre les systèmes eubactériens et archaées, ce dernier étant lui-même très similaire au système eucaryote. Le génie génétique actuel ne peut imiter ce phénomène qu'à renfort de modification des structures des gènes des sélénoprotéines sujettes au transfert entre les domaines pour leur étude,<sup>[I.201, I.218]</sup> même si des possibles passerelles ont été récemment trouvées.<sup>[I.219, I.220]</sup>

Une corrélation a également pu être établie entre le mode de vie des organismes bactériens et leur capacité à utiliser le sélénium sous forme de sélénocystéine (possession des gènes *selABCD*) et/ou de 2-sélénouridine (possession des gènes *selD* et *ybbB*). Les résultats sont illustrés par les 3 histogrammes à droite de la figure 1.36, page ci-contre.

La tendance statistique observée est une utilisation de la 2-sélénouridine chez les bactéries vivant à des températures basse et ambiante, contrastant avec l'utilisation de la sélénocystéine chez les bactéries vivant à des températures plus hautes.



**Figure 1.36** – Étude de la capacité d'utilisation de la sélénocystéine chez les eubactéries. [I.165]  
 [à gauche] Corrélation entre la distance phylogénique des organismes et l'identité de séquence des protéines SelA, SelB et SelD. Chaque graphique contient 100 points, correspondant à 100 comparaisons de paires d'organismes sélectionnées au hasard. Un cas de transfert horizontal de gène est marqué par une flèche. Le coefficient de corrélation CC entre parenthèse est celui obtenu après correction (suppression des cas de transferts de gènes).  
 [à droite] Corrélation entre le mode de vie d'un organisme (anaérobie/aérobie, thermophile/psychrophile) et sa capacité à utiliser le sélénium, en Sec et/ou en Se<sup>2</sup>U. "Sec+" et "Sec-" signifient la capacité et l'incapacité, respectivement, à utiliser la sélénocystéine.

La corrélation entre l'exposition à l'oxygène et l'utilisation du sélénium conduit à une conclusion identique à celle obtenue chez les eucaryotes : les organismes anaérobies utilisent statistiquement plus la sélénocystéine que les organismes aérobies, alors que l'utilisation de la 2-séléno-uridine est indépendante de ce paramètre environnemental. L'exemple des protéines homologues au domaine HesB, ayant un rôle dans la biogenèse des clusters fer-soufre catalytiques, illustre bien la tendance de l'utilisation des sélénoprotéines : les organismes anaérobies ont statistiquement plus conservés la version sélénié, alors qu'aucun organisme aérobie ne possède cette version.

Les conclusions de ces études sont que les facteurs environnementaux ont une forte influence sur le maintien ou la perte de la capacité à utiliser la sélénocystéine, comme suggéré par les études effectuées sur les eucaryotes. Le gain, le maintien ou la perte de cette capacité sont néanmoins des processus fréquents chez les eubactéries, et l'évolution à travers les âges de cette capacité est dynamique et réversible. Le fragile équilibre entre les avantages et les inconvénients supposés de l'utilisation de cet acide aminé réactif explique cette dynamique et l'usage limité qui en est fait dans le vivant. En revanche, aucune piste supplémentaire concernant l'émergence de l'utilisation de la sélénocystéine ne peut être tirée de ces données, mise à part la confirmation d'une origine ancestrale de cette capacité pour les trois domaines du vivant.

Des questions subsistent donc au sujet de l'émergence de l'utilisation de la sélénocystéine et l'évolution indépendante de cette capacité dans les trois domaines du vivant.

---

L'étude de la machinerie de biosynthèse et d'insertion de la sélénocystéine est un domaine de choix pour chercher les réponses aux interrogations laissées par cet acide aminé, et le manque de données sur ce système chez les eubactéries demande davantage d'investigations.

La version bactérienne de cette machinerie est un objet d'étude naturel pour notre équipe, dont les travaux sont axés autour de la compréhension des détails de la traduction bactérienne et de ses exceptions.

Cette thématique a donc été abordée dans deux directions, se concentrant tout d'abord sur le système bactérien d'insertion de la sélénocystéine, sujet d'étude de la thèse soutenue en juillet 2007 par Nicolas Soler,<sup>[I.221]</sup> puis sur le système de biosynthèse bactérien de cet acide aminé sur son ARN<sub>t</sub>, sujet du présent travail.

# Chapitre 2

## Les acteurs de la synthèse de la sélénocystéine sur son ARN<sub>t</sub>

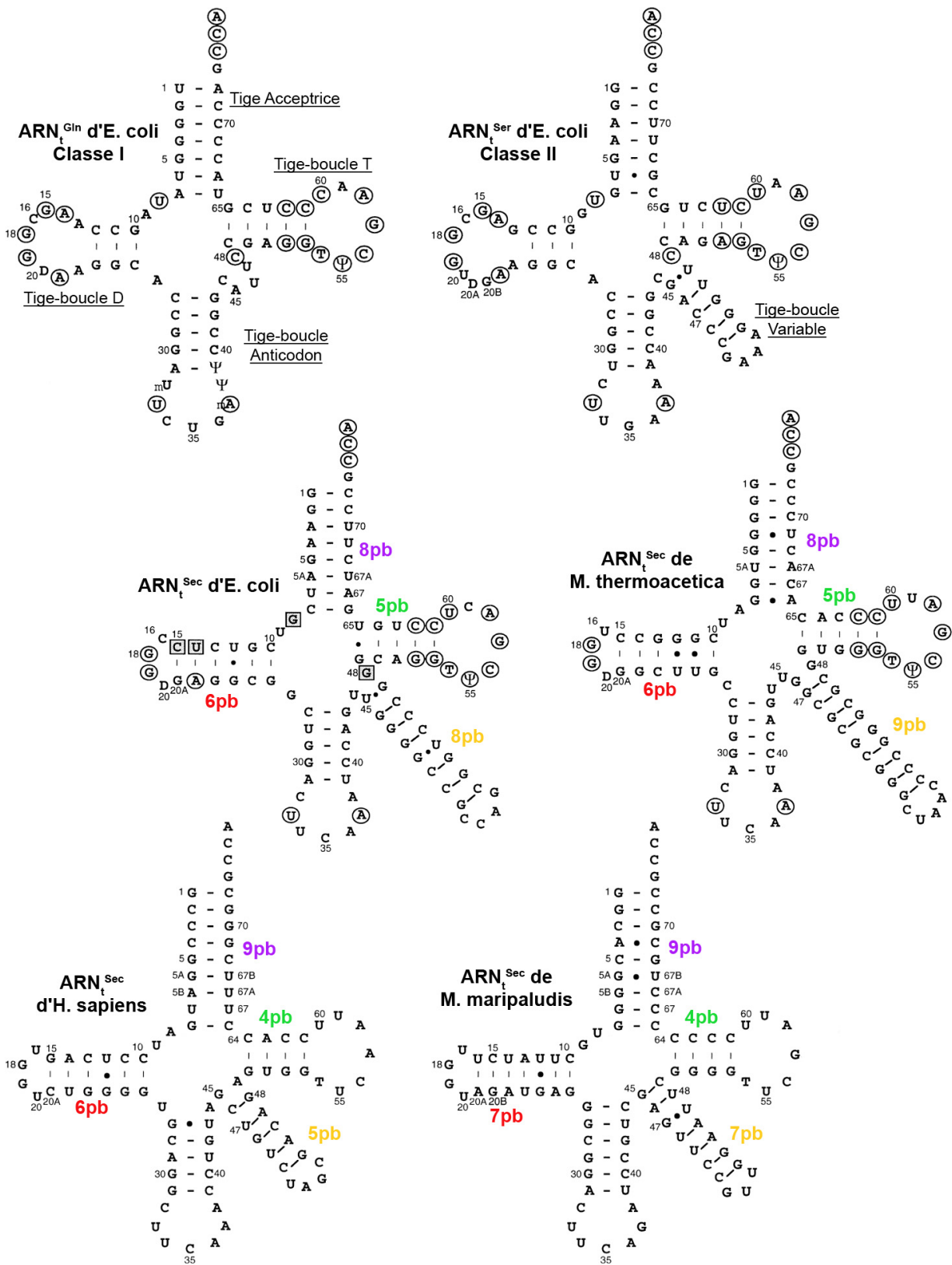
### Sommaire

---

<b>2.1</b>	<b>L'ARN<sub>t</sub><sup>Sec</sup></b> . . . . .	<b>81</b>
2.1.1	Structures et déterminants . . . . .	81
2.1.2	L'importance de la modification des bases . . . . .	85
2.1.3	Évolution et lien entre les ARN <sub>t</sub> <sup>Sec</sup> . . . . .	86
<b>2.2</b>	<b>Le sélénophosphate SePO<sub>3</sub><sup>3-</sup></b> . . . . .	<b>87</b>
2.2.1	Présentation et historique . . . . .	87
2.2.2	Propriétés chimiques . . . . .	88
2.2.3	La sélénophosphate synthétase . . . . .	89
<b>2.3</b>	<b>La sélénocystéine synthase SelA</b> . . . . .	<b>90</b>
2.3.1	Présentation . . . . .	90
2.3.2	Dans la famille des enzymes PLP-dépendantes, je demande...	92
2.3.2.1	Présentation du cofacteur . . . . .	92
2.3.2.2	Étude évolutionnelle et comparaison structurale . . . . .	92
2.3.3	Mécanisme . . . . .	94
2.3.4	Stœchiométrie et discrimination chez SelA . . . . .	94
2.3.4.1	Une question d'encombrement ? . . . . .	94
2.3.4.2	Coopérativité entre les sites . . . . .	96
2.3.4.3	Discrimination vis-à-vis des substrats . . . . .	97
2.3.5	Quel lien évolutionnel avec les enzymes homologues? . . . . .	98
2.3.5.1	Comparaison avec les SepCysS . . . . .	99
2.3.5.2	Le lien logique avec les SepSecS . . . . .	100
2.3.5.3	La similarité des "SelA-like" . . . . .	104
2.3.6	Rares sont les structures décameriques . . . . .	106
<b>2.4</b>	<b>Conclusion</b> . . . . .	<b>107</b>

---





**Figure 2.1** – Structures secondaires des ARN<sub>t</sub><sup>Sec</sup> d'*Escherichia coli* et de *Moorella thermoacetica* comparés à des ARN<sub>t</sub> canoniques bactériens de Classe I et II (au-dessus) et à des ARN<sub>t</sub><sup>Sec</sup> eucaryote et archae (en-dessous). Les bases cerclées sont invariantes ou semi-invariantes parmi les ARN<sub>t</sub> canoniques bactériens. Des carrés grisés indiquent la déviation par rapport à ce consensus de l'ARN<sub>t</sub><sup>Sec</sup> d'*E. coli*. Les ARN<sub>t</sub><sup>Sec</sup> eucaryote et archae ne portent pas l'indication de leurs bases modifiées.

Une meilleure connaissance du système de biosynthèse de la sélénocystéine chez les eubactéries pourrait permettre d'accéder à des informations supplémentaires sur les questions évolutives sous-jacentes à l'utilisation de la sélénocystéine dans le vivant.

Nous avons donc choisi d'étudier le détail structural et enzymatique des acteurs de cette biosynthèse du Sec-ARN<sub>t</sub><sup>Sec</sup> à partir du Ser-ARN<sub>t</sub><sup>Sec</sup> et de sélénophosphate, effectuée par la sélénocystéine synthase SelA.

Ce chapitre présente toutes les données disponibles sur ces différents acteurs, sur lesquelles se base notre étude. Nos travaux ont eu pour objet deux représentants célèbres du monde eubactérien, *Escherichia coli* et *Moorella thermoacetica* (anciennement *Clostridium thermoaceticum*). Les données concernant la machinerie de biosynthèse de la sélénocystéine de ces organismes seront abordées plus en détail.

## 2.1 L'ARN<sub>t</sub><sup>Sec</sup>

L'ARN<sub>t</sub><sup>Sec</sup> est l'élément pivot dans le mécanisme de synthèse et d'incorporation de la sélénocystéine dans les protéines : il est présent à toutes les étapes, du précurseur à l'incorporation co-traductive de la sélénocystéine.

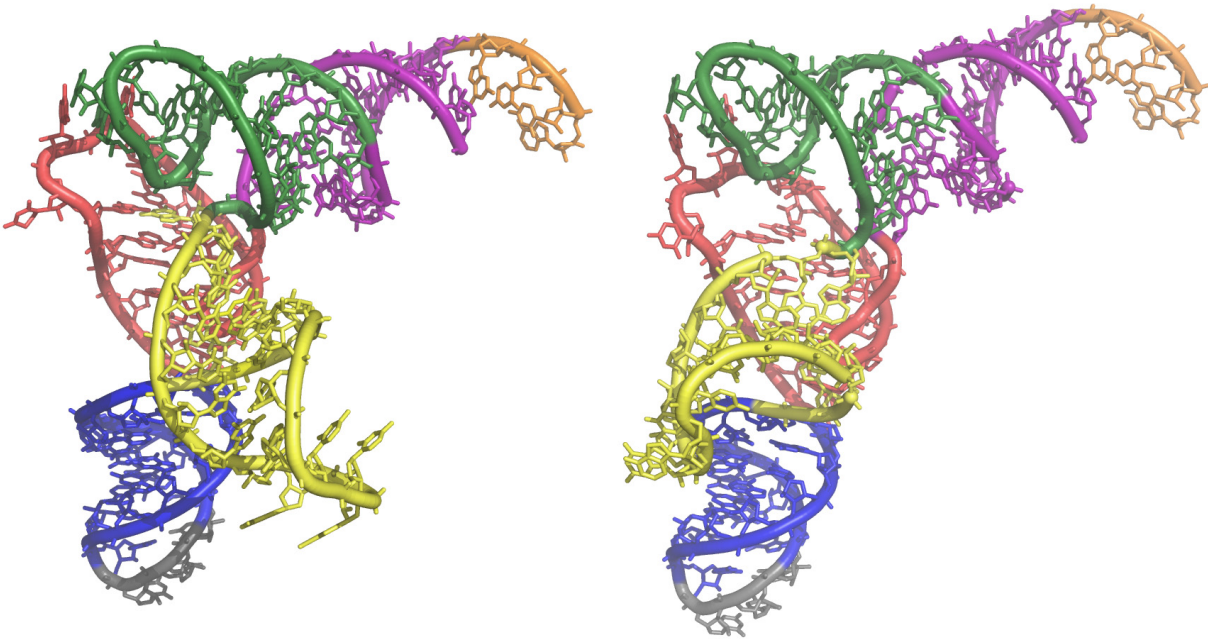
Dans le système bactérien, cet ARN<sub>t</sub><sup>Sec</sup> doit donc interagir successivement avec la séryl-ARN<sub>t</sub> synthétase (SerRS), la sélénocystéine synthase (SelA), puis avec son facteur d'élongation spécifique SelB. La structure de cet ARN<sub>t</sub><sup>Sec</sup> est donc caractérisée par des déterminants permettant à ces enzymes de le reconnaître, et permettant à celles qui ne doivent pas interagir avec lui, comme le facteur d'élongation classique EF-Tu, de le discriminer dans la population des ARN<sub>t</sub> cellulaires.

### 2.1.1 Structures et déterminants

Par rapport aux ARN<sub>t</sub> canoniques de Classe I et II, les ARN<sub>t</sub><sup>Sec</sup> bactériens ont comme caractéristiques particulières<sup>[1,222]</sup> (voir schéma 2.1, page précédente) :

- ◇ Une tige acceptrice de 8 paires de bases (pb) au lieu de 7. La numérotation utilisée est celle de C. Baron, J. Heider et A. Böck,<sup>[1,223]</sup> avec une paire numérotée 5A-67A. (Une 1<sup>ère</sup> numérotation, utilisée par A. Schon, A. Böck *et al.*,<sup>[1,222]</sup> comptait une paire 3A-69A.)
- ◇ Une tige D de 6pb au lieu de 3 ou 4, et une boucle D qui en conséquence ne contient que 4 nucléotides.
- ◇ Une tige-boucle variable d'une longueur plus importante que pour n'importe quel ARN<sub>t</sub> de la Classe II : l'ARN<sub>t</sub><sup>Ser</sup> à anticodon GCU n'atteint que 7pb chez d'*E. coli*.
- ◇ Une déviation par rapport aux nucléotides consensus retrouvés à certaines positions dans tous les ARN<sub>t</sub> canoniques (cercés sur la schéma 2.1). En particulier, les nucléotides G8, U14, C15 et G48 sont caractéristiques de l'ARN<sub>t</sub><sup>Sec</sup> d'*E. coli*.
- ◇ Un anticodon UCA complémentaire au codon UGA de l'ARN<sub>m</sub>.

En comparaison, les  $ARN_t^{Sec}$  d'eucaryotes et d'archaes possèdent ces mêmes caractéristiques, à l'exception d'une tige acceptrice de 9pb, compensée par une tige T de seulement 4pb.<sup>[I.224,I.225]</sup> La structure tridimensionnelle des  $ARN_t^{Sec}$  des trois domaines du vivant sont similaires, leur bras accepteur composé de la tige acceptrice et de la tige-boucle T comptant 13pb, contre 12 pour la plupart des  $ARN_t$  canoniques.



**Figure 2.2** – Structures tridimensionnelle des  $ARN_t^{Sec}$  procaryote (gauche) et eucaryote (droite), obtenues par expériences d'empreintes chimiques et enzymatiques et par modélisation.<sup>[I.119,I.224,I.226]</sup>

La tige-boucle variable, en jaune, est flexible et peut donc prendre plusieurs orientations. Le code couleur respecte celui du schéma 1.4, page 38.

Il est d'ailleurs remarquable que ces  $ARN_t^{Sec}$  structurellement similaires soient interchangeables entre espèces : les  $ARN_t^{Sec}$  de *Moorella thermoacetica*,<sup>[I.227]</sup> de *Desulfomicrobium baculatum*<sup>[I.227]</sup> et des eucaryotes *Homo sapiens*<sup>[I.228]</sup> et *Xenopus laevis*<sup>[I.228]</sup> supportent la synthèse et l'insertion de la sélénocystéine dans une souche d'*Escherichia coli* chez laquelle le gène *selC* a été supprimé.

## Les déterminants de reconnaissance de l' $ARN_t^{Sec}$

Cette interchangeabilité implique que les déterminants utiles à la distinction de l' $ARN_t^{Sec}$  dans la cellule, lors des différentes étapes du système aboutissant à la traduction des sélénoprotéines, sont conservés entre les espèces.

De nombreuses études ont permis de découvrir quels sont les déterminants spécifiques à chaque étape de ce processus de biosynthèse et d'insertion de la sélénocystéine, généralement chez le modèle *Escherichia coli*.

## Reconnaissance par SerRS

La première étape de la biosynthèse de la sélénocystéine est l'estérification d'une sérine sur l'ARN<sub>t</sub><sup>Sec</sup> par la séryl-ARN<sub>t</sub> synthétase, qui accomplit également cette réaction sur le pool des ARN<sub>t</sub><sup>Ser</sup> cellulaires.

Comme un point de forte similarité entre les ARN<sub>t</sub><sup>Sec</sup> et ARN<sub>t</sub><sup>Ser</sup> est leur tige-boucle variable de taille importante (voir figure 2.1, page 80), il fut supposé que ce caractère est ce qui permet la qualification de ces ARN<sub>t</sub> comme substrat de l'enzyme. La structure de SerRS complexée avec un ARN<sub>t</sub><sup>Ser</sup> (voir figure 1.7, page 42) permet d'ailleurs de visualiser le mode de reconnaissance de cette enzyme : les contacts protéine/ARN se font à la base de la tige-boucle variable et au site d'estérification.

L'ARN<sub>t</sub><sup>Sec</sup> n'est pas un très bon substrat de réaction pour SerRS : la réaction avec l'ARN<sub>t</sub><sup>Sec</sup> est 100 fois moins efficace qu'avec un ARN<sub>t</sub><sup>Ser</sup> chez *E. coli*, cette différence tombant à un facteur 10 dans le système d'*H. sapiens*.<sup>[I.229]</sup> De façon surprenante, l'insertion de la tige-boucle variable d'un ARN<sub>t</sub><sup>Ser</sup> à la place de la tige-boucle variable de l'ARN<sub>t</sub><sup>Sec</sup> ne permet pas d'obtenir un ARN<sub>t</sub> chimère réagissant plus efficacement (3<sup>ème</sup> ligne) :

ARN <sub>t</sub>	K <sub>m</sub>	K <sub>cat</sub>	Efficacité relative
ARN <sub>t</sub> <sup>Ser</sup> 1 anticodon CGA	1,2 μM	0,11 s <sup>-1</sup>	100
ARN <sub>t</sub> <sup>Sec</sup> natif	11,2 μM	0,011 s <sup>-1</sup>	1,1
ARN <sub>t</sub> <sup>Sec</sup> (TBvariable d'ARN <sub>t</sub> <sup>Ser</sup> )	4,6 μM	0,001 s <sup>-1</sup>	0,2
ARN <sub>t</sub> <sup>Sec</sup> (ΔA5a-U67a)	9,4 μM	0,023 s <sup>-1</sup>	2,8

**Tableau 2.1** – Valeurs caractéristiques de la cinétique enzymatique de SerRS d'*Escherichia coli* selon l'ARN<sub>t</sub> substrat.<sup>[I.230]</sup> L'ARN<sub>t</sub><sup>Sec</sup>(ΔA5a-U67a) est un mutant avec délétion des nucléotides A5a-U67a, conduisant à une tige acceptrice de 7pb.

Donc, si le site de reconnaissance utilisé par l'enzyme est bien la tige-boucle variable de ces ARN<sub>t</sub>, l'efficacité de la réaction d'estérification est un subtil équilibre entre la longueur de la tige acceptrice et le positionnement du bras variable dans le site de l'enzyme, équilibre qui détermine la composition de la tige-boucle variable de l'ARN<sub>t</sub><sup>Sec</sup>.<sup>[I.230]</sup>

Comme les caractéristiques nécessaires à cet équilibre sont dépendantes de la SerRS et donc de l'organisme considéré, il n'est pas étonnant que les ARN<sub>t</sub><sup>Sec</sup> des eucaryotes *Homo sapiens* et *Xenopus laevis* soient des substrats très inefficaces pour la réaction chez *Escherichia coli*.<sup>[I.228]</sup>

## Reconnaissance par Sela

La deuxième étape de la biosynthèse bactérienne de la sélénocystéine est le remplacement de l'hydroxyle de la sérine par un sélénium, effectué par la sélénocystéine synthase Sela grâce à du sélénophosphate. Cette enzyme doit donc discerner les séryl-ARN<sub>t</sub><sup>Sec</sup> des séryl-ARN<sub>t</sub><sup>Ser</sup> présents dans la cellule.

De façon non surprenante, le mutant ARN<sub>t</sub><sup>Sec</sup>(TBvariable d'ARN<sub>t</sub><sup>Ser</sup>) est sélénié aussi efficacement que l'ARN<sub>t</sub><sup>Sec</sup> natif.<sup>[I.230]</sup> Par contre, le mutant ARN<sub>t</sub><sup>Sec</sup>(ΔA5a-U67a) est sélénié très lentement par Sela, montrant l'importance de la longueur de la tige acceptrice pour l'efficacité de la réaction.

Les  $ARN_t^{Sec}$  eucaryotes sont, eux, assez efficacement sélénés par SelA.<sup>[1.228]</sup> Cette étape de la biosynthèse de la sélénocystéine n'est donc pas très discriminatoire vis-à-vis du substrat de réaction, si tant est que celui-ci possède une structure adaptée.

### Reconnaissance par SelB

La reconnaissance du sélénocystéinyl- $ARN_t^{Sec}$  par SelB est essentielle pour l'insertion co-traductrice de cet acide aminé lors de la synthèse de sélénoprotéines.

Pour que cette reconnaissance ait lieu, deux éléments sont nécessaires. L'acide aminé estérifié doit être une sélénocystéine ; les séryl- $ARN_t^{Sec}$  ne se complexent pas avec SelB. Le bras accepteur doit avoir les 13pb caractéristiques de l' $ARN_t^{Sec}$  : le mutant  $ARN_t^{Sec}(\Delta A5a-U67a)$  ne permet pas l'insertion co-traductrice de sélénocystéine.<sup>[1.230]</sup> Par contre, le mutant  $ARN_t^{Sec}(TBvariable\ d'ARN_t^{Ser})$  et les  $ARN_t^{Sec}$  eucaryotes<sup>[1.228]</sup> permettent l'incorporation de la sélénocystéine chez *Escherichia coli*.

Le discriminant utilisé par SelB est donc la longueur de la tige acceptrice de l' $ARN_t^{Sec}$ .

### Discrimination par EF-Tu

Pour garder l'ambiguïté du sens du codon UGA, il est important que le facteur d'élongation de la traduction n'interagisse pas avec les  $ARN_t^{Sec}$ .

Effectivement, l'affinité d'EF-Tu pour les  $ARN_t^{Sec}$  est 1000 fois plus faible que pour les autres  $ARN_t$  cellulaires.<sup>[1.231]</sup>

L'élément le plus discriminant pour EF-Tu paraît, là aussi, être la longueur inhabituelle de la tige acceptrice : le mutant  $ARN_t^{Sec}(\Delta A5a-U67a)$  interagit avec EF-Tu, qu'il ait une sérine ou une sélénocystéine estérifiée à son extrémité 3'.<sup>[1.230]</sup>

Pourtant, J. Rudinger *et al.*<sup>[1.232]</sup> montrèrent que d'autres éléments sont en cause : un  $ARN_t^{Asp}$  de levure ayant également un bras accepteur de 13pb est un substrat pour le facteur d'élongation EF-Tu de *Thermus thermophilus*, donc la longueur du bras n'est pas un discriminant suffisant. L'élément discriminant est apparu après une étude par mutagenèse sur l' $ARN_t^{Sec}$  d'*E. coli* : il s'agit des 8<sup>ième</sup>, 9<sup>ième</sup> et 10<sup>ième</sup> paires de bases du bras accepteur, correspondant chez les bactéries à la dernière paire de bases de la tige acceptrice et des deux premières de la tige-boucle T. Il s'agit pour *E. coli* du motif C7-G66/G49-U65/C50-G64 qui est absent de tout les  $ARN_t$  canoniques procaryotes connus.

### Importance des autres éléments d'identité

Il apparaît donc que les éléments les plus discriminants sont présents sur le bras accepteur. La tige-boucle variable joue également un rôle dans la reconnaissance de l' $ARN_t^{Sec}$  par la SerRS.

Les autres éléments qui différencient l' $ARN_t^{Sec}$  des  $ARN_t$  canoniques n'ont-ils pas leur importance lors d'une étape de la biosynthèse et l'incorporation de la sélénocystéine ?

◇ Parmi les déviations par rapport aux nucléotides consensus retrouvés sur tous les  $ARN_t$ , la paire inhabituelle G11-C24 peut être inversée en C11-G24 sans aucune conséquence sur

la réactivité de l'ARN<sub>t</sub><sup>Sec</sup> obtenu.<sup>[I.223]</sup> La purine présente en position 8 peut également être mutée en l'uridine U8 universellement conservée sans que cela affecte la fonction de l'ARN<sub>t</sub><sup>Sec</sup>.<sup>[I.223]</sup>

◇ Par contre, le remplacement des uridines U9 et U14, respectivement par G9 ou A14, diminue l'efficacité globale de l'insertion de la sélénocystéine, mais les ARN<sub>t</sub><sup>Sec</sup> obtenus restent fonctionnels.<sup>[I.223]</sup>

◇ L'anticodon UCA de l'ARN<sub>t</sub><sup>Sec</sup> ne semble servir qu'à l'interaction avec l'ARN<sub>m</sub> : le changement de cet anticodon par celui d'un ARN<sub>t</sub><sup>Cys</sup> (GCA) permet tout de même la biosynthèse de la sélénocystéine et l'insertion de cet acide aminé au codon UGA, avec une efficacité de 30% par rapport à l'ARN<sub>t</sub><sup>Sec</sup> natif. Lorsque l'anticodon est modifié à une autre position que la position "wobble", la biosynthèse et l'insertion d'une sélénocystéine restent efficaces à condition de changer le codon de l'ARN<sub>m</sub> de la sélénoprotéine en conséquence.<sup>[I.223]</sup>

◇ Les ARN<sub>t</sub><sup>Sec</sup> eucaryotes et archaïques subissent une étape supplémentaire dans la synthèse de la sélénocystéine : le séryl-ARN<sub>t</sub><sup>Sec</sup> est phosphorylé par PSTK avant de réagir avec la sélénocystéine synthase SepSecS. Pour cette réaction, les archaïques ne semblent pas utiliser le même élément de discrimination que les eucaryotes<sup>[I.233]</sup> : la PSTK eucaryote utilise la spécificité de la tige-boucle D des ARN<sub>t</sub><sup>Sec</sup>,<sup>[I.234]</sup> alors que la PSTK archaïque reconnaît des éléments de la tige acceptrice, en particulier les paires de bases 2-71, 3-70 et 5-68.<sup>[I.233]</sup>

Au final, l'utilité des éléments qui différencient l'ARN<sub>t</sub><sup>Sec</sup> des autres ARN<sub>t</sub>, si elle ne réside pas dans la discrimination directe lors de réactions enzymatiques, se traduit par une potentielle importance structurale : ils permettent à l'ARN<sub>t</sub><sup>Sec</sup> d'adopter la conformation nécessaire à ces réactions.

Un alignement des séquences des ARN<sub>t</sub><sup>Sec</sup> de *M. thermoacetica* et d'*E. coli*, et des isoaccepteurs ARN<sub>t</sub><sup>Ser</sup> d'*E. coli* est présenté au § 4.7, page 187, sur lequel sont résumées des différentes positions de discrimination des ARN<sub>t</sub><sup>Sec</sup>.

## 2.1.2 L'importance de la modification des bases

Bien que peu de bases de l'ARN<sub>t</sub><sup>Sec</sup> soient modifiées chez les procaryotes, ces modifications ont une importance réactionnelle, et un ARN<sub>t</sub><sup>Sec</sup> synthétisé par transcription *in vitro* ne présente pas la même réactivité que les ARN<sub>t</sub><sup>Sec</sup> obtenus *in vivo* et cela quelle que soit l'étape de synthèse ou d'incorporation de la sélénocystéine.<sup>[I.228]</sup>

Ces modifications ont une importance encore plus grande chez les eucaryotes, où certaines sont utilisées à des fins de régulation de l'expression des sélénoprotéines. Les modifications de bases présentes chez la plupart des eucaryotes<sup>[I.235]</sup> sont les suivantes :

◇ Une pseudo-uridine ( $\Psi$ ) en position 55, et une 1-méthyladénosine (m1A) en position 58.

◊ Une N6-isopentényladénosine (i6A), présente en position A37. L'absence de cette modification inhibe la synthèse de sélénoprotéines, de façon protéine- et organe-spécifique chez les mammifères, sans pour autant empêcher la biosynthèse de la sélénocystéine sur l' $ARN_t^{Sec}$ .<sup>[I.236,I.237]</sup> L'absence de cette modification paraît altérer la maturation de l' $ARN_t^{Sec}$ .

◊ Une 5-méthylcarboxyméthyluridine (mcm5U), présente à la position U34. La modification de l'U34 se produit après la modification des autres bases de l' $ARN_t^{Sec}$ , dans le cytoplasme.<sup>[I.238]</sup> Cette uridine U34, en position “wobble”, est ensuite méthylée par SECp43 en 5-méthylcarboxyméthyl-2'-O-méthyluridine (mcm5Um) lors de la maturation finale du Sec- $ARN_t^{Sec}$ .<sup>[I.135]</sup> La présence dans les cellules eucaryotes de deux populations d' $ARN_t^{Sec}$ , en proportion séléo-dépendante, avait déjà été remarquée en 1991.<sup>[I.239]</sup> Cette méthylation est un moyen de régulation de l'expression des sélénoprotéines, l' $ARN_t^{Sec}$  qui contient une mcm5Um34 étant l'espèce utilisée pour l'insertion de la sélénocystéine lors de leur synthèse.<sup>[I.240]</sup> En cas de mutation empêchant cette méthylation, seules quelques sélénoprotéines vitales comme les thiorédoxine réductases (TrxR) continuent d'être synthétisées à un niveau quasi-normal.<sup>[I.136]</sup> Cette maturation est gouvernée par la structure primaire et tertiaire de l' $ARN_t^{Sec}$ , qui doit posséder toutes les modifications de bases précédentes.<sup>[I.241]</sup> De plus, la structure tridimensionnelle de l' $ARN_t^{Sec}$  est affectée par la présence de ce méthyle en 2' du ribose de l'uridine occupant la position “wobble”.

Il apparaît donc que les modifications de l' $ARN_t^{Sec}$  lui permettent d'avoir la structure requise pour les différentes étapes de biosynthèse et d'incorporation de la sélénocystéine. Elles peuvent également être utilisées comme moyen de régulation comme le montre les conséquences de la maturation finale de l' $ARN_t^{Sec}$  par méthylation.

Cette dernière caractéristique différencie les systèmes bactérien et eucaryote par la complexité plus importante du second. Pourtant, comme les  $ARN_t^{Sec}$  sont inter-échangeables entre espèces et entre domaines du vivant, ils sont une preuve supplémentaire de l'origine partagée des systèmes eubactériens et eucaryotes concernant la sélénocystéine.

### 2.1.3 Évolution et lien entre les $ARN_t^{Sec}$

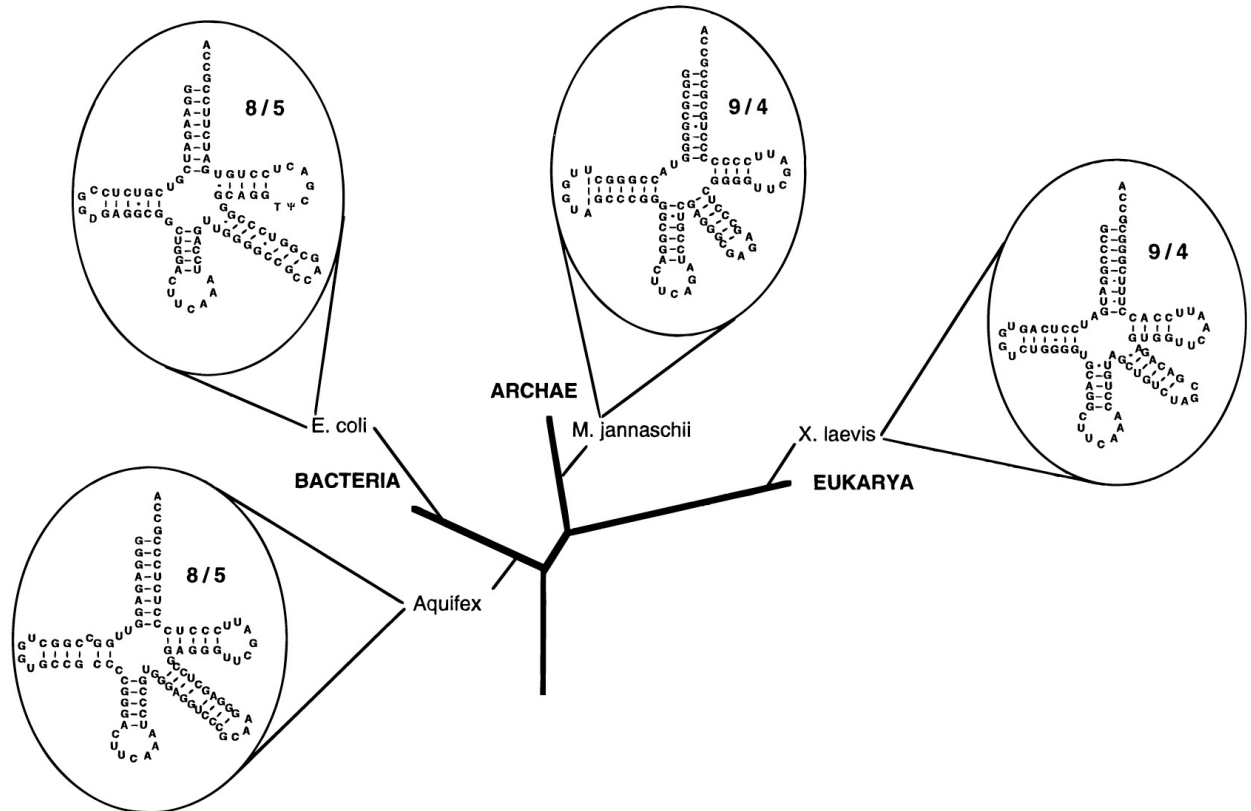
L' $ARN_t^{Sec}$  est une structure conservée dans le vivant, ce qui suggère son utilisation primordiale déjà pour le décodage du codon UGA en sélénocystéine.

Comme la séquence du gène *selC* est trop courte pour tenter de déterminer une phylogénie précise par comparaison de sa composition entre espèces, les seules observations peuvent être qualitatives. Ces observations montrent que les divergences des  $ARN_t^{Sec}$  entre espèces suivent la distance entre les ARN ribosomaux de ces espèces.<sup>[I.227]</sup>

Le point principal permettant de considérer que les éléments de structure et de discrimination de l' $ARN_t^{Sec}$  ont été conservés au cours de l'évolution est la possibilité pour un  $ARN_t^{Sec}$  de supporter la synthèse de sélénoprotéines dans un organisme exogène et éloigné.<sup>[I.227,I.228]</sup>

Le mécanisme de synthèse et ses acteurs peuvent donc également être considérés comme évolutionnairement liés.<sup>[I.242]</sup> Dans cette logique, l' $ARN_t^{Sec}$  le plus divergeant de

ses congénères connus est celui de *Chlamydomonas reinhardtii*, qui confirme une origine commune aux  $\text{ARN}_t^{\text{Sec}}$  des trois domaines du vivant, toutes branches confondues.<sup>[1.243]</sup>



**Figure 2.3** – Conservation de la structure générale des  $\text{ARN}_t^{\text{Sec}}$  au cours de l'évolution, malgré le passage des  $\text{ARN}_t^{\text{Sec}}$  eucaryotes et archaées à une forme 9pb/4pb (longueur des tiges acceptrices et T, respectivement) (Tiré de N. Hubert *et al.*<sup>[1.224]</sup>).

## 2.2 Le sélénophosphate $\text{SePO}_3^{3-}$

Pour que la réaction de biosynthèse de la sélénocystéine soit possible, les sélénocystéine synthases bactériennes (SelA) et eucaryotes (SepSecS) ont besoin d'un donneur de sélénium approprié. Cette molécule est le sélénophosphate, qui prend part au complexe ternaire de transfert du sélénium vers l'acide aminé estérifié sur l' $\text{ARN}_t^{\text{Sec}}$ .

### 2.2.1 Présentation et historique

Peu après la découverte des gènes nécessaires à l'incorporation de la sélénocystéine chez *Escherichia coli*, la protéine SelD fut caractérisée comme l'activateur de sélénium en vu de son transfert lors des réactions de synthèse de la sélénocystéine et de la sélénouridine<sup>[1.244]</sup> (voir § 4, page 31 au sujet de la sélénouridine). L'identité de l'espèce séléniée activée fut connue grâce à des études par RMN du phosphore  $^{31}\text{P}$  de ce composé, qui ont

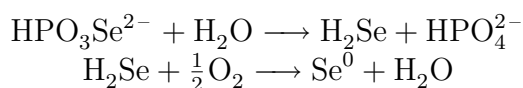


permis de découvrir la liaison entre les atomes de sélénium et de phosphore,<sup>[I.95]</sup> puis de découvrir sa véritable identité par comparaison avec du sélénophosphate synthétique.<sup>[I.245]</sup>

Pour son étude, le sélénophosphate peut être obtenu de deux façons différentes<sup>[I.246]</sup> : soit de façon enzymatique à partir d'ATP et de sélénure HSe<sup>-</sup> par l'action de sélénophosphate synthétase purifiée, soit chimiquement par hydrolyse de triméthylsilylsélénophosphate (TMSO)<sub>3</sub>PSe. Dans tous les cas, la réaction doit être effectuée de façon anaérobie, à cause de la grande sensibilité du composé.<sup>[I.245-I.247]</sup>

## 2.2.2 Propriétés chimiques

En effet, la manipulation du sélénophosphate demande l'utilisation de solvants dégazés et en atmosphère sans oxygène : soit sous azote, soit sous argon. Le sélénophosphate est très sensible à l'air, le sélénium le composant étant labile par réaction avec l'oxygène.



selon T. Mizutani *et al.*<sup>[I.247]</sup>

La réaction de décomposition du sélénophosphate en milieu aqueux oxygéné produit du phosphate et du sélénium élémentaire reconnaissable par la couleur rouge du précipité obtenu. La stabilité du sélénophosphate a été étudiée par RMN du phosphore <sup>31</sup>P.<sup>[I.247]</sup> Pour une solution aqueuse de sélénophosphate, il en ressort les résultats suivants :

Conditions de conservation			Demi-vie
Température	Atmosphère	pH	
-80°C	sous azote	6,8	∞
0°C	sous azote	6,8	840h
-19°C	air	6,8	740h
22°C	sous azote	4,0	240h
22°C	sous azote	6,8	33h
0°C	air	6,8	32h
0°C	sous azote	9,0	9h

**Tableau 2.2** – Demi-vie de SePO<sub>3</sub><sup>3-</sup> dans un tampon Hepes 0,2 M, avec 20 mM de MgCl<sub>2</sub> et 20 mM de KCl, selon les conditions de stockage.<sup>[I.247]</sup>

L'acide sélénophosphorique H<sub>3</sub>PO<sub>3</sub>Se est un triacide, dont le pKa2 et pKa3 sont de 4,6 et 8,8 respectivement.<sup>[I.246]</sup> Les données du tableau 2.2 permettent de conclure que la stabilité des espèces acides est donc remarquablement plus importante que celle de la base SePO<sub>3</sub><sup>3-</sup>. Les conditions privilégiées de stabilité du sélénophosphate sont donc une température basse, dans un tampon acide et anoxique.

La stabilité en milieu cellulaire peut être expliquée par le potentiel réducteur du cytosol.<sup>[I.247]</sup> Le taux de sélénophosphate dans le corps humain est de 0,2 mg par kg de poids corporel, ce qui est bien loin de la dose toxique, la DL50 ayant été mesurée à 37,5 mg/kg

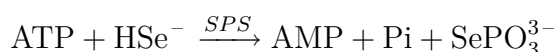
corporel chez la souris.<sup>[I.247]</sup> Ce taux corporel est sans doute régulé par l'activité de la sélénophosphate synthétase.

### 2.2.3 La sélénophosphate synthétase

La sélénophosphate synthétase est le produit du gène *selD* chez les eubactéries, soit une protéine de 347 acides aminés et de 36,7 kDa pour *Escherichia coli*.<sup>[I.244]</sup>

L'enzyme remplissant cette fonction chez les eucaryotes est SPS2, une enzyme homologue de SelD. L'autre homologue eucaryote de SelD est SPS1 dont la spécificité a été abordée au § 1.3.3, page 49. Cette enzyme n'utilise pas le sélénium comme substrat, mais produirait du sélénophosphate à partir de sélénocystéine libre.<sup>[I.144]</sup>

SelD et SPS2 catalyse la réaction suivante :



Le mécanisme de cette réaction a été abordé en détails au § 1.5.2, page 65 et la position de cette réaction dans le cycle de synthèse de la sélénocystéine est visible sur le schéma 1.8, page 42.

SelD est spécifique à ces substrats, sélénium et ATP : la mise en présence d'ions sulfure ne permet pas d'obtenir de thiophosphate, et les autres nucléosides-triphosphates ne réagissent pas avec l'enzyme.<sup>[I.209]</sup> Les  $K_m$  respectifs ont été déterminés à 0,9 mM pour l'ATP et 46  $\mu\text{M}$  pour le sélénium. Cette dernière valeur est au-dessus de ce qui peut être considéré comme biologiquement possible, les milieux de culture ne contenant pas plus de 0,1-1  $\mu\text{M}$  de sélénium, mais malgré de nouvelles mesures dans des conditions strictement anaérobies et réductrices, une valeur encore importante de 7,3  $\mu\text{M}$  fut obtenue.<sup>[I.209]</sup> Cela alimente l'hypothèse de l'existence d'une espèce sélénée pour le transfert du sélénium réduit dans la cellule, sans doute sous la forme d'une protéine de transport, comme exposé au § 1.3.3, page 47 et par la figure 1.14, page 48, et qui fournirait SelD en substrat.

Les résidus de l'enzyme SelD d'*Escherichia coli* indispensables à son activité catalytique sont la cystéine Cys17 et la lysine Lys20.<sup>[I.248]</sup> Certaines sélénophosphate synthétases bactériennes sont des sélénoprotéines, dans lesquelles la sélénocystéine est d'ailleurs présente à la position équivalente à la cystéine Cys17 de SelD d'*Escherichia coli*, mais il a été vu au § 1.5.2, page 65 que ces sélénio-enzymes ne sont pas plus efficaces que leurs homologues thiolées (voir tableau 1.5, page 66).

Pour pouvoir catalyser la réaction, l'enzyme nécessite la présence d'ions magnésium  $\text{Mg}^{2+}$ , d'autres ions divalents, comme  $\text{Mn}^{2+}$  ou  $\text{Co}^{2+}$ , ne pouvant remplir son rôle.<sup>[I.249]</sup> Un ion monovalent, différent de  $\text{Na}^+$  ou  $\text{Li}^+$ , est également requis : un ion potassium  $\text{K}^+$  (ou dans une moindre mesure  $\text{NH}_4^+$  ou  $\text{Rb}^+$ ) peut remplir ce rôle.<sup>[I.209]</sup>

Les sélénophosphate synthétases ont été étudiées avec intérêt, à cause de leur double rôle de fournisseur de sélénium actif dans les voies de synthèse de la sélénocystéine et la sélénouridine, et leur possible rôle dans d'autres métabolismes.<sup>[I.102]</sup> Les études phylogéniques sur les différents acteurs des utilisations du sélénium montrent que SelD, SPS1 et SPS2 sont reliées phylogénitiquement, et suivent généralement les branches classiques des divergences évolutives.<sup>[I.91]</sup>

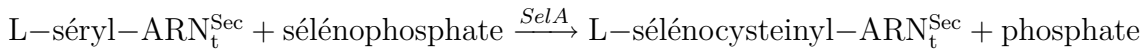
## 2.3 La sélénocystéine synthase SelA

La sélénocystéine synthase SelA est une protéine exclusivement procaryotique, qui est l'objet d'étude central de ce travail. Les sélénocystéines synthases eucaryotes et archeaes, appelées SepSecS, seront abordées au § 2.3.5.2, page 100, et utilisées comme point de comparaison avec SelA.

### 2.3.1 Présentation

« Sélénocystéine synthase » est le nom usuel qui définit la classe d'enzyme EC 2.9.1.-. et SelA définit la classe EC 2.9.1.1 des « L-séryl-ARN<sub>t</sub><sup>Sec</sup> sélénium transférase ».

Cette enzyme catalyse la réaction suivante :



Il s'agit du produit du gène *selA* découvert lors de la caractérisation des gènes nécessaires à l'incorporation du sélénium chez *Escherichia coli* en 1988<sup>[1.3]</sup> (voir l'historique § 1.1, page 35). Le gène *selA* forme une unité transcriptionnelle avec *selB*, se trouvant à la minute 80 du chromosome d'*Escherichia coli*. Cette organisation en opéron se retrouve chez un certain nombre de procaryotes possédant la machinerie d'incorporation de la sélénocystéine (l'organisation des opérons est visible en Annexe A.2, page 229) montrant le lien fort entre ces deux protéines, également essentielles dans cette machinerie et ayant évolué de façon lié.

Pourtant, parmi les gènes mis à jour et leurs différents produits, ARN ou protéines, SelA est l'enfant pauvre de la famille. En effet, peu de travaux ont été effectués sur cette protéine : seules 7 publications<sup>[1.250-1.256]</sup> ont eu pour sujet d'étude SelA, dont les premières en 1991, soit trois ans après sa découverte. Ces premières études permettent toutefois d'accéder à d'importantes informations sur SelA d'*Escherichia coli*.<sup>[1.250,1.251]</sup> Les études sur SelA de *Moorella thermoacetica* viendront plus tard.<sup>[1.255]</sup>

Organisme	<i>Escherichia coli</i>	<i>Moorella thermoacetica</i>
Nombre de résidus	463	470
Masse monomérique	50607 Da	50590 Da
pI théorique	6,05	6,75

**Tableau 2.3** – Tableau récapitulatif des données concernant SelA.

Ainsi, pour les organismes *E. coli* et *M. thermoacetica*, SelA est composé respectivement de 463 et 470 acides aminés (Séquences en Annexes A.4, page 232), ce qui en fait une protéine de 50607 et 50590 Daltons, respectivement.

En solution, la protéine se présente sous forme décamerique, et est éluée avec une masse perçue d'approximativement 600000 Da en chromatographie sur gel d'exclusion stérique. L'analyse de la séquence de cette protéine montre une forte hydrophobicité, et un pI théorique de 6,05 et 6,75 respectivement pour *E. coli* et *M. thermoacetica*.

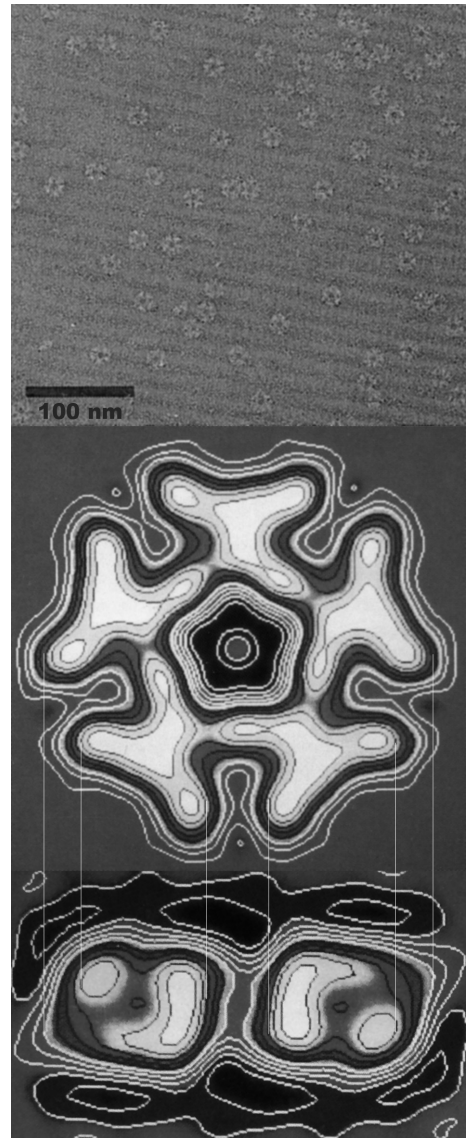
Bien qu'aucune structure de SelA n'ait été déterminée jusqu'à présent, nous avons tout de même une idée sur l'assemblage de ce décamer, grâce aux travaux de Engelhardt *et al.*<sup>[1.254]</sup> qui ont mis au point des conditions d'obtention de cliché de la protéine d'*Escherichia coli* par cryo-microscopie électronique. Comme des clichés ont été obtenus dans deux orientations, de "face" et de "profil", ils permettent de caractériser le décamer de «pentamère de dimère». En effet, les données de volume et de masse des sous-unités sont en accord avec la symétrie d'ordre 2 des clichés de "profil", et permettent d'affirmer que les extensions en forme de T sont formées de deux monomères entrelacés.<sup>[1.254]</sup>

Une étude similaire a été effectuée sur SelA de *Moorella thermoacetica*,<sup>[1.256]</sup> et montre que la protéine se présente en solution sous la forme d'un décamer de dimension similaire à celui d'*Escherichia coli*. L'avancée des techniques et la mise au point des méthodes permettraient même d'obtenir, à terme, une structure tridimensionnelle à basse résolution de la protéine, par cryomicroscopie électronique.

Par contre, bien que la forme du décamer montre une symétrie axiale d'ordre 2 et une symétrie rotationnelle d'ordre 5, ces images ne permettent pas de conclure quant à la symétrie ou l'asymétrie des rôles des monomères dans chaque dimère et dans le décamer, même si chacun possède un site actif fonctionnel.

L'existence d'un site actif fonctionnel par monomère a été montrée par la présence d'autant de molécules de cofacteurs que de monomères.<sup>[1.250]</sup>

Un décamer comporte dix molécules de pyridoxal-5'-phosphate (PLP), ce qui fait de SelA un membre de la grande famille des enzymes PLP-dépendantes. Ce cofacteur est lié de façon covalente à la protéine par l'intermédiaire de la lysine K295 pour *Escherichia coli* et la lysine K292 pour *Moorella thermoacetica*.<sup>[1.255]</sup>



**Figure 2.4** – Cliché de cryomicroscopie électronique, projection de face et projection de profil.<sup>[1.254]</sup> La dimension du décamer est de 190Å/190Å/65Å avec un trou central de 40Å.

## 2.3.2 Dans la famille des enzymes PLP-dépendantes, je demande...

### 2.3.2.1 Présentation du cofacteur

Le pyridoxal phosphate, qui est un dérivé de la vitamine B6, est sans doute le cofacteur organique le plus commun et versatile, utilisé par tous les organismes et impliqué dans 4% des activités enzymatiques référencées.<sup>[I.257]</sup> Chez les procaryotes, 1,5% des gènes codent pour des enzymes PLP-dépendantes.

Toutes ces enzymes, à l'exception des glycogène phosphorylases, participent à des voies métaboliques impliquant, parmi d'autres composés aminés, une majorité d'acides aminés.

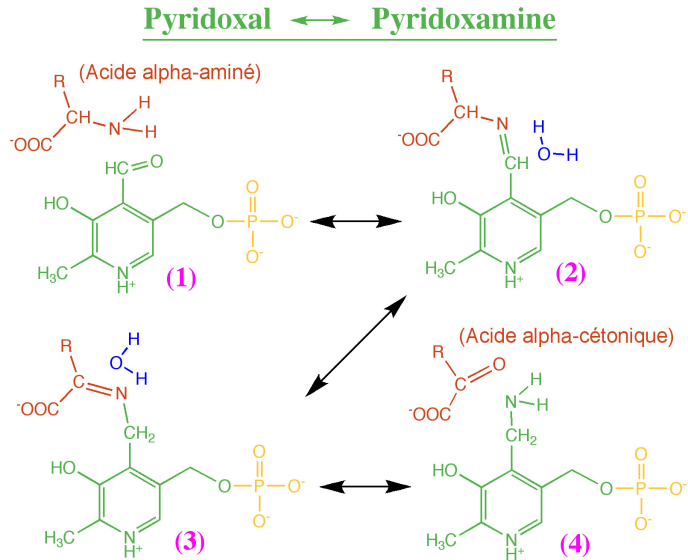
Les enzymes PLP-dépendantes qui ont pour substrats des amino-acides permettent des réactions de transfert du groupement amine, de décarboxylation, d'interconversion entre acides aminés de la série L et de la série D, et d'élimination ou de remplacement de groupes chimiques sur les carbones  $\gamma$  ou  $\beta$ , ce qui est le cas pour la sélénocystéine synthase.

### 2.3.2.2 Étude évolutionnelle et comparaison structurale

Toutes les enzymes PLP-dépendantes peuvent se répartir en 5 groupes structurels, ayant chacun une origine évolutionnelle commune à en juger par l'homologie importante caractérisant les enzymes d'un groupe. La famille structurale la plus importante est celle comptant parmi ses membres la sélénocystéine synthase SelA : il s'agit des enzymes PLP-dépendantes de type structural I ou «Fold-Type I». Cette classification structurale remplace l'ancienne classification dans laquelle SelA était un membre de la famille  $\alpha$ , incluse dans la super-famille  $\alpha/\gamma$ .

Ces classifications, créées pour regrouper les différentes enzymes PLP-dépendantes, permettent en général de trouver quels sont les membres ayant le plus de caractéristiques communes, pour faciliter les études par homologie. Malheureusement, quelle que soit la classification utilisée, la place de SelA et l'identité des membres évolutionnellement proches n'ont pu être déterminées sans ambiguïté. En effet, les diverses tentatives<sup>[I.255, I.258-I.260]</sup> se soldent par des résultats différents. Les arbres phylogéniques correspondants sont présents en Annexe A.5, page 234.

Ceci doit être lié à la particularité de la sélénocystéine synthase d'être la seule enzyme PLP-dépendante, avec la SepCysS décrite précédemment (voir § 1.4.3, page 55), à avoir



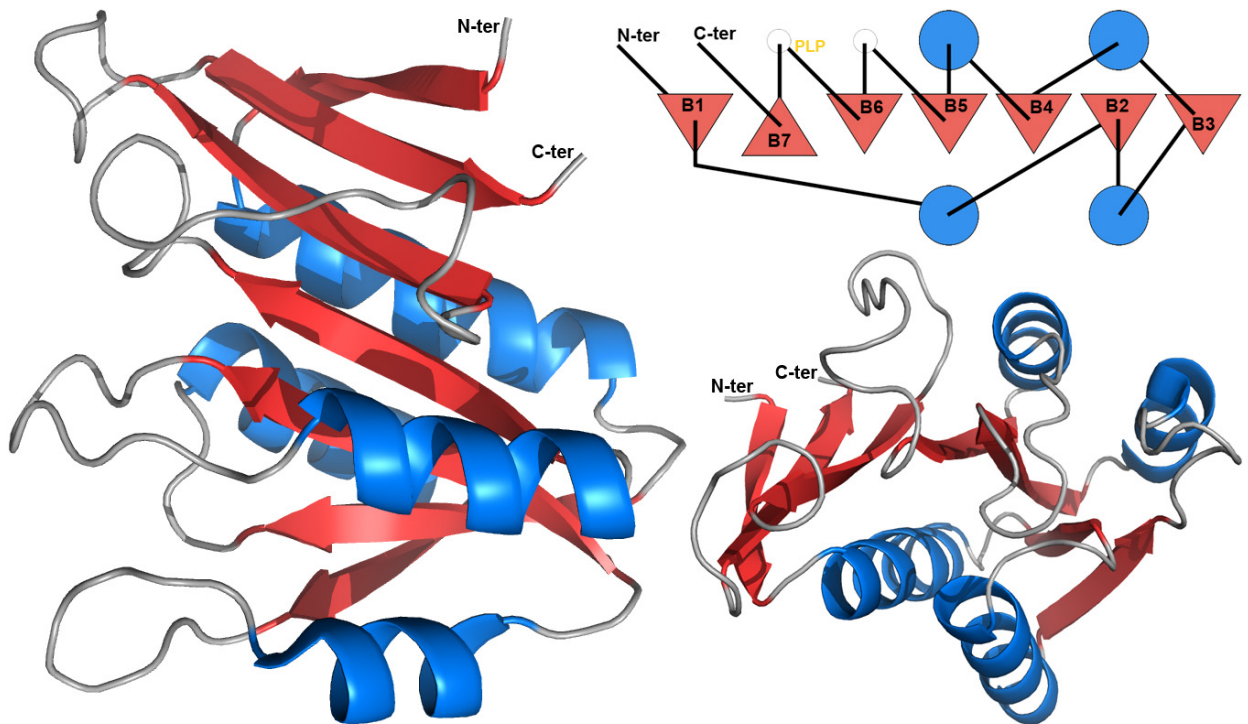
**Figure 2.5** – Schéma présentant les différentes formes du cofacteur Vitamine B6 ou PLP. (1)Pyridoxal-phosphate et acide  $\alpha$ -aminé, (2)formation d'une base de Schiff par déshydratation, (3)forme tautomère, (4)Pyridoxamine-phosphate et acide  $\alpha$ -cétonique.

comme substrat un acide aminé estérifié sur un  $\text{ARN}_t$ , et non un acide aminé libre. En effet, les autres enzymes de modification d'acide-aminé sur leur  $\text{ARN}_t$  connues, que sont les amidotransférases (AdT), ne sont pas des enzymes PLP-dépendantes. Ces enzymes ont été décrites au § 1.4.1, page 49 et permettent la synthèse par transamidation des glutaminyl- $\text{ARN}_t^{\text{Gln}}$  et asparaginylyl- $\text{ARN}_t^{\text{Asn}}$ .

### Comparaison structurale

Toutes les enzymes PLP-dépendantes de type structural I ou «Fold-Type I» sont caractérisées par certains éléments structuraux très conservés malgré une identité de séquence faible. Elles possèdent toutes des régions dont la structure est conservée et des régions de contacts hydrophobiques conservés, en particulier au niveau du “domaine” coeur portant le pyridoxal phosphate.<sup>[1.261]</sup>

Ce coeur est structuré en feuillet  $\beta$  composé de 7 brins ( $\beta_1$ - $\beta_7$ - $\beta_6$ - $\beta_5$ - $\beta_4$ - $\beta_2$ - $\beta_3$ ), pris dans un sandwich  $\alpha/\beta/\alpha$  caractéristique. Ce feuillet  $\beta$  est parallèle, à l'exception du brin  $\beta_7$ , anti-parallèle. Le cofacteur pyridoxal phosphate est porté par la connexion entre les brins  $\beta_6$  et  $\beta_7$ .



**Figure 2.6** – Structure tridimensionnelle du coeur conservé des enzymes PLP-dépendantes, en sandwich  $\alpha/\beta/\alpha$ .

Les notations “N-ter” et “C-ter” représentent les connexions avec les domaines N-terminal et C-terminal de la protéine. Le coeur présenté ici provient de la structure de la protéine IscS d'*Escherichia coli* (n° PDB : 1P3W).

D'autres motifs structuraux peuvent compléter ce coeur, au niveau des boucles entre les brins  $\beta$ , selon la complexité des enzymes PLP-dépendantes considérées. Les protéines

PLP-dépendantes ont aussi une organisation oligomérique en commun : elles se présentent sous forme de dimères, voire de tétramères, et rarement sous forme d'hexamères. Sela est le seul exemple d'enzyme PLP-dépendante décamerique connu. Des exemples de structures tridimensionnelles complètes sont disponibles en Annexe A.6, page 235, pour SepCysS et pour d'autres enzymes PLP-dépendantes impliquées dans le métabolisme du sélénium, comme IscS ou CsdB, ainsi que pour d'autres représentantes du «Fold-Type I».

L'appartenance de la sélénocystéine synthase à la famille des enzymes PLP-dépendantes ne permet pas d'avoir des informations évolutionnelles à son sujet, mais permet d'avoir une idée de la structure de son domaine catalytique, par homologie structurale avec les autres enzymes de «Fold-Type I». De plus, les mécanismes réactionnels mis en jeu par le cofacteur pyridoxal phosphate sont bien connus, ce qui a permis de postuler une séquence catalytique pour le fonctionnement de l'enzyme.

### 2.3.3 Mécanisme

Le cofacteur pyridoxal phosphate permet des réaction très différentes, mais passant par des intermédiaires similaires. Les mécanismes réactionnels impliquant les substrats acides-aminés ont été longuement étudiés,<sup>[I.258,I.262–I.264]</sup> ce qui permet de connaître le cas particulier d'une  $\beta$ -substitution, réaction catalysée par la sélénocystéine synthase.

Certaines observations expérimentales ont permis à K. Forchhammer et A. Böck<sup>[I.251]</sup> d'affiner ce mécanisme général pour le cas de la sélénocystéine synthase. En particulier, en l'absence du donneur de sélénium, le complexe Sela / séryl-ARN<sub>t</sub><sup>Sec</sup> produit est stable et purifiable sur chromatographie par gel d'exclusion stérique. La décomposition de ce complexe stable en présence d'hydruure forme de l'alanine, alors que du pyruvate est formé par décomposition lente en l'absence de tout réactif. Ceci montre qu'après la création de la liaison covalente entre le PLP et la sérine chargé sur l'ARN<sub>t</sub><sup>Sec</sup>, cette dernière est déshydratée pour obtenir un intermédiaire aminoacrylyl-ARN<sub>t</sub><sup>Sec</sup>, espèce (2) du schéma réactionnel 2.7, page ci-contre.

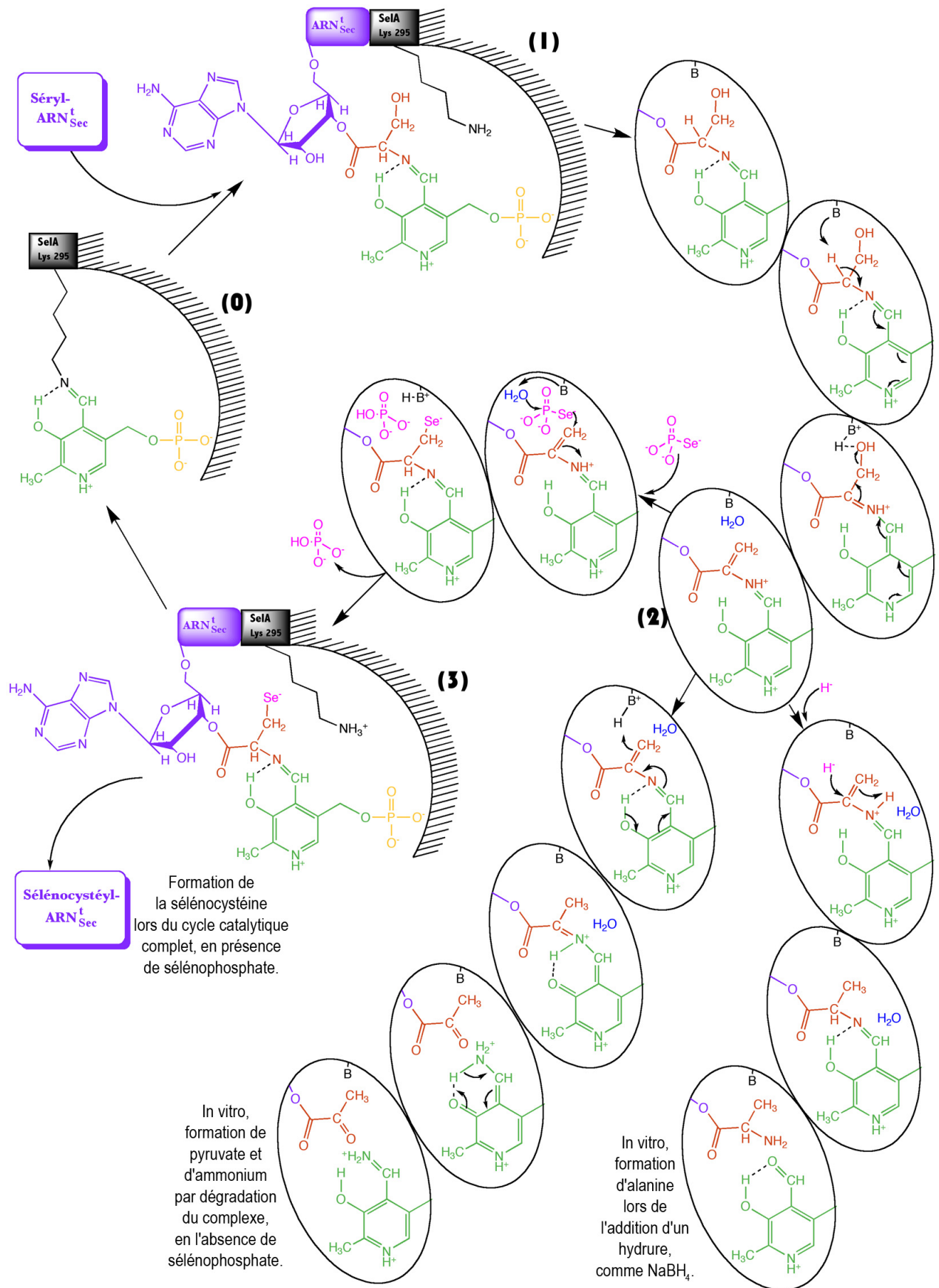
Lorsque la source de sélénium est disponible, identifiée comme étant une molécule de sélénophosphate, l'addition du sélénium sur la double liaison de l'intermédiaire acrylyl-ARN<sub>t</sub><sup>Sec</sup> permet la formation du sélénocystéinyl-ARN<sub>t</sub><sup>Sec</sup>, en (3) sur le schéma 2.7.

L'ensemble de ces réactions, qui se résume dans le nom « L-séryl-ARN<sub>t</sub><sup>Sec</sup> sélénium transférase », se passe donc au site actif contenant le cofacteur pyridoxal phosphate. Or la protéine forme un décamerère en solution, et chaque monomère possède un site actif intégrant un cofacteur PLP. Il pourrait donc y avoir dix sites fonctionnels sur ce décamerère.

### 2.3.4 Stœchiométrie et discrimination chez Sela

#### 2.3.4.1 Une question d'encombrement ?

Nous venons de voir ce qui se passe au site actif lors de la réaction. Mais les dix sites actifs du décamerère sont-ils tous fonctionnels ? La symétrie spatiale de la macromolécule permet de penser que les dix sites sont indiscernables sur le décamerère sans substrat.



**Figure 2.7** – Mécanisme enzymatique de formation de la sélénocystéine par Sela.

(0) Sela et son cofacteur, (1) Complexation du Sérilyl-ARN<sub>t</sub><sup>Sec</sup>, (2) Forme intermédiaire stable : l'aminocrylyl-ARN<sub>t</sub><sup>Sec</sup>, (3) Complexe avant relargage du sélénocystéinyl-ARN<sub>t</sub><sup>Sec</sup>.

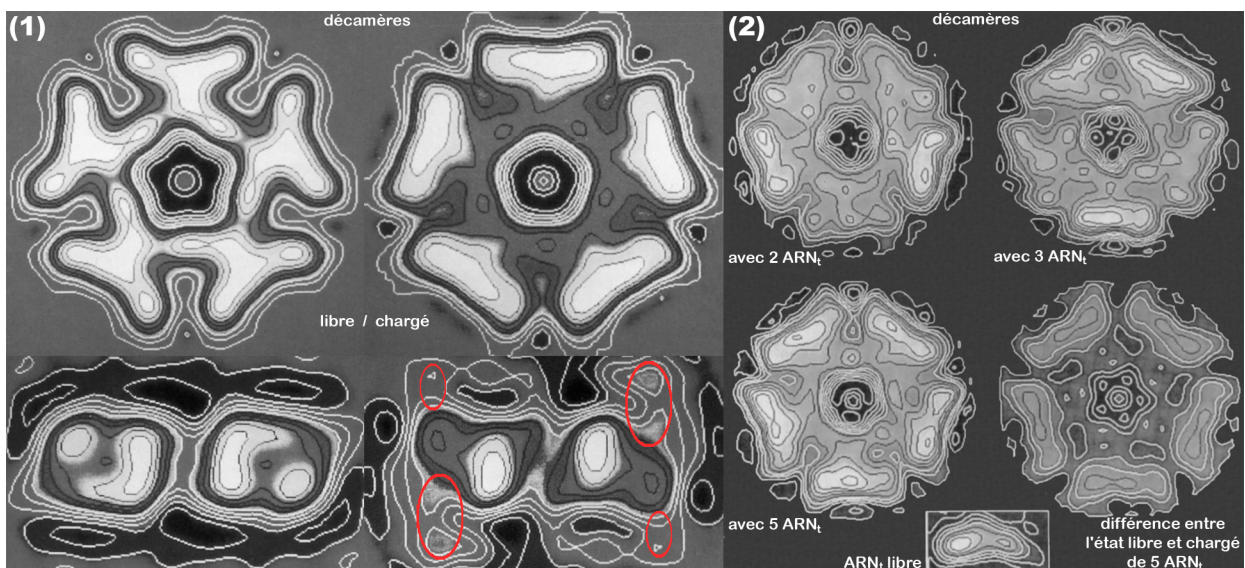
La base "B" représentée dans les encarts elliptiques pourrait être la Lys295 ou une autre base présente sur l'enzyme.



Et comme les complexes Sela / séryl- $ARN_t^{Sec}$  sont stables, il serait théoriquement possible d'obtenir l'état de complexation de 1 décimère de Sela pour 10 séryl- $ARN_t^{Sec}$ .

Pourtant, seul l'état maximal de 5 séryl- $ARN_t^{Sec}$  pour 1 décimère de Sela a été observé,<sup>[I.251]</sup> soit expérimentalement un rapport Sela / séryl- $ARN_t^{Sec}$  de 2,1:1 (ou 2:0,95 soit 95% de dimères chargés) dans des conditions d'excès en solution de 1:3.

La confirmation de cette stœchiométrie vint avec les clichés du complexe obtenu par cryo-microscopie électronique en 1992.<sup>[I.254]</sup> Lors d'une charge partielle du décimère de Sela, les différents états suivants ont été observés : 2, 3 ou 5 séryl- $ARN_t^{Sec}$  sur 1 décimère. Cette charge partielle du décimère a été obtenue dans des conditions d'excès en solution de 1:1 (Sela / séryl- $ARN_t^{Sec}$ ).



**Figure 2.8** – Projections de clichés de cryo-microscopie électronique du complexe.<sup>[I.254]</sup>  
 (1) Comparaison entre l'enzyme libre (à gauche) et l'enzyme chargée de 5 séryl- $ARN_t^{Sec}$  (à droite). (2) Différents états de charge du complexe, et différence entre les états 0 et 5. En encart, cliché de l' $ARN_t^{Sec}$  seul, échelle respectée (75Å/30Å).

La position de la masse supplémentaire présente sur le complexe indique que l' $ARN_t^{Sec}$  interagit avec, au maximum, les deux monomères formant le T de la sous-unité dimérique. L'étude conclut que les molécules de séryl- $ARN_t^{Sec}$  se situent indifféremment sur le "dessus" ou le "dessous" du décimère, mais qu'une sous-unité dimérique ne pouvait interagir qu'avec une seule molécule de séryl- $ARN_t^{Sec}$ . Cette conclusion est obtenue par observation du cliché du décimère de profil avec  $ARN_t^{Sec}$  (chargé), où il semblerait que soit visible une augmentation de densité sur les 2 dimères visibles et la rangée des 2 dimères suivants, avec une configuration de complexation de l' $ARN_t^{Sec}$  suivant une schématique alternée "au-dessus" / "en-dessous"<sup>[I.254]</sup>, indiquée par les ellipses rouges sur la figure 2.8.

### 2.3.4.2 Coopérativité entre les sites

Donc, grâce aux précédents clichés obtenus en cryomicroscopie électronique, l'équipe de A. Böck émet l'hypothèse d'une coopérativité négative<sup>[I.254]</sup> entre les deux sites présents

sur une sous-unité dimérique : lorsqu'un complexe est formé entre une molécule de séryl-ARN<sub>t</sub><sup>Sec</sup> et un dimère, par un site acif et une face de ce dimère, le deuxième site actif et l'autre face du dimère ne peuvent former, simultanément, une autre liaison avec un autre partenaire nucléique. On peut donc parler de modulation homotrope au sein d'un dimère.

L'équipe conclue également que l'absence de tendance préférentielle vers un état très chargé (ou peu chargé) du décamère, mais une existence d'états intermédiaires observables, tend à réfuter la possibilité d'une forte coopérativité entre les dimères. Il est tout de même intéressant de noter que les états intermédiaires observés sont des remplissages des positions 1 & 3 pour un décamère doublement chargé, et 1, 3 & 5 pour un décamère triplement chargé (nous considérons ici que la position de départ et le sens de la numérotation des sous-unités sont permutable, dû aux équivalences induites par les symétries de la protéine).

### 2.3.4.3 Discrimination vis-à-vis des substrats

#### Reconnaissance de l'ARN<sub>t</sub> substrat

Les études menées sur l'entité ARN<sub>t</sub> nous permettent de connaître les propriétés le différenciant des autres ARN<sub>t</sub> présents dans la cellule (voir § 2.1, page 81). Elles nous ont également permis de connaître les discriminants utilisés par les enzymes responsables de chaque étape de la synthèse et de l'incorporation de la sélénocystéine.

Il ressortait de ces études que la protéine SelA n'est pas très discriminatoire, vu que les ARN<sub>t</sub><sup>Sec</sup> d'autres espèces ou mutés sont substrats de la réaction, avec une efficacité tout de même assez faible pour le mutant ARN<sub>t</sub><sup>Sec</sup>(ΔA5a-U67a) à tige acceptrice réduite d'une paire de base. C'est la structure spécifique de l'ARN<sub>t</sub><sup>Sec</sup> qui est donc reconnue.

Qu'apportent à ce sujet les études menées sur SelA ?

K. Forchhammer *et al.*<sup>[1.251]</sup> ont confirmé que SelA différencie le séryl-ARN<sub>t</sub><sup>Sec</sup> du séryl-ARN<sub>t</sub><sup>Ser</sup>, mais ont pu montrer que l'ARN<sub>t</sub><sup>Sec</sup> non chargé en sérine interagit également avec SelA, sans toutefois former de complexe covalent stable. De plus, aucune réaction n'est observée entre SelA et de la sérine libre : la présence de l'ARN est donc un prérequis à la réaction.

Le mécanisme de reconnaissance du substrat provient bien de la spécificité de l'ARN<sub>t</sub><sup>Sec</sup>, qui est acteur de l'interaction avec SelA et permet la réaction de son séryl-ester avec le cofacteur PLP du site actif.

Il pourrait être envisagé que le site actif soit rendu accessible pour la réaction par un changement conformationnel local dû à l'interaction de l'enzyme avec l'ARN. Cette hypothèse est confortée par des études biochimiques<sup>[1.253]</sup> montrant une protection importante de ce site actif : l'échange continu du tampon dans lequel se trouve l'enzyme, par dialyse, ne permet pas de lui faire perdre son cofacteur PLP, et le cyanure de potassium KCN, qui inactive habituellement ce cofacteur, ne provoque qu'une inhibition partielle de la réaction. À l'inverse, la liaison aldimine entre l'enzyme et le PLP est réduite par traitement de SelA avec de l'hydrure NaBH<sub>4</sub>, ce qui ne modifie pas son interaction avec l'ARN<sub>t</sub><sup>Sec</sup> non chargé en sérine, mais la liaison définitive du cofacteur à l'enzyme empêche la formation d'un complexe avec une molécule de séryl-ARN<sub>t</sub><sup>Sec</sup>. K. Forchhammer *et al.*<sup>[1.253]</sup> en

concluent l'existence d'une coopérativité entre le site actif contenant le PLP et le site de fixation de l'ARN.

Bien que les sites de reconnaissance entre l'enzyme et l'ARN<sub>t</sub><sup>Sec</sup> ne soient pas connus, différentes hypothèses à confirmer existent : une reconnaissance de la tige acceptrice de l'ARN<sub>t</sub><sup>Sec</sup> ou même une interaction de l'anticodon seraient possibles, même si les mutations de celui-ci n'ont aucun effet sur l'activité de SelA (voir § 2.1.1, page 83).

Globalement, la double reconnaissance spécifique du séryl-ARN<sub>t</sub><sup>Sec</sup> par SelA et du sélénocystéinyl-ARN<sub>t</sub><sup>Sec</sup> par SelB, lors des étapes de synthèse et d'insertion de la sélénocystéine, sont comparables à la haute spécificité de sélection des acides aminés canoniques par le couple formé des aminoacyl-ARN<sub>t</sub> synthétases et du facteur d'élongation EF-Tu.

### Discrimination du donneur de sélénium

Pour que la réaction d'obtention du sélénocystéinyl-ARN<sub>t</sub><sup>Sec</sup> soit efficace, le groupe hydroxyle de la sérine substrat ne doit pouvoir être remplacé que par du sélénium. Or nous avons vu que le soufre et le sélénium sont chimiquement suffisamment proches pour ne pas être discernables dans certains mécanismes biologiques (voir § 2, page 20). Le donneur de sélénium a été identifié<sup>[1.96,1.245]</sup> comme étant du sélénophosphate  $\text{SePO}_3^{3-}$  synthétisé par l'enzyme SelD. Or du thiophosphate présent dans la cellule pourrait réagir au site actif de SelA à la place du sélénophosphate, et former ainsi du cystéinyl-ARN<sub>t</sub><sup>Sec</sup>. Des mesures ont été effectuées par P. Tormay *et al.*<sup>[1.255]</sup> lors d'expériences de cinétique enzymatique sur SelA, donnant les résultats suivants :

Substrat	$Km$ $\mu\text{M}$	$kcat$ $\text{min}^{-1}$	$kcat/Km$ $\text{min}^{-1} \cdot \mu\text{M}$
Sélénophosphate*	0,3	1,3	4,3
Thiophosphate	4,0	0,052	0,013

**Tableau 2.4** – Spécificité de SelA pour les donneurs de chalcogènes.<sup>[1.255]</sup> \*(Séléniure + ATP + SelD)

Donc, dans la réaction de  $\beta$ -remplacement opérée par SelA, le mécanisme employé permet de différencier le sélénophosphate du thiophosphate, avec un rapport de spécificité de plus de 300 pour 1. Cette étude a également permis de montrer que d'autres formes du sélénium ou du soufre, comme du sélénite  $\text{HSe}^-$  ou du sulfite  $\text{HS}^-$ , ne permettent pas d'obtenir de produit, sauf pour des concentrations importantes, sans pertinence biologique (supérieures à 1 mM).

SelA est donc assez spécifique, tant vis à vis de son ARN<sub>t</sub> substrat que du donneur de sélénium.

### 2.3.5 Quel lien évolutif avec les enzymes homologues ?

Trois types d'enzymes homologues à SelA, dont des structures tridimensionnelles ont été obtenues, peuvent être comparées à la sélénocystéine synthase bactérienne. L'intérêt de cette comparaison et de l'étude des relations phylogéniques entre ces enzymes est d'obtenir

des informations structurales par homologie et, si cela est possible, de construire un modèle structural de SelA.

Les trois types d'enzymes considérées sont les SepCysS archaïques, les SepSecS eucaryotes et archaïques, et les protéines archaïques "SelA-like".

La comparaison de SelA et de ces enzymes posent autant de difficultés qu'elle présente d'intérêts. En effet, aucune de ces protéines ne se présente sous forme de décamère, mais par contre elles font toutes partie de la famille des enzymes PLP-dépendantes de «Fold-Type I» et, pour les SepCysS et SepSecS, catalysent la transformation d'un acide aminé estérifié sur un ARN<sub>t</sub>. La réaction catalysée par les protéines "SelA-like" archaïques n'est pas connue, mais il a été montré qu'elles n'ont pas d'activité sélénocystéine synthase, malgré l'identité de séquence qu'elles ont avec SelA.<sup>[1.265]</sup>

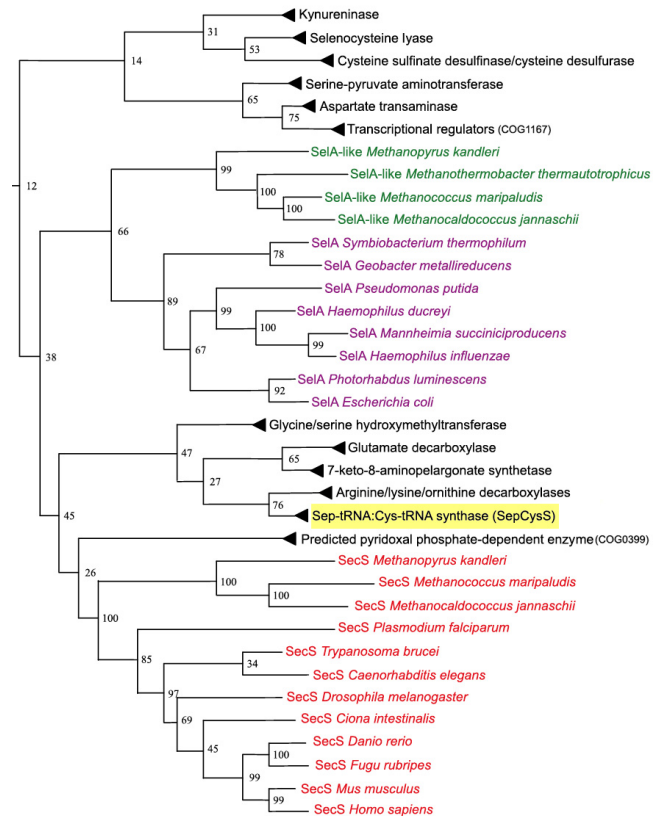
La difficulté la plus importante reste le fait que toutes ces protéines ne sont présentes que chez les eucaryotes et archaïques, et donc que s'il paraît possible qu'elles aient un lien évolutif, quel qu'il soit, avec la sélénocystéine synthase bactérienne SelA, les divergences entre les domaines du vivant n'ont pu que renforcer les divergences de séquences et de structures entre ces enzymes.

### 2.3.5.1 Comparaison avec les SepCysS

Comme lors de la synthèse de la sélénocystéine chez les eubactéries, la synthèse indirecte de la cystéine chez certaines archaïques se fait en deux étapes : une més-estérification par une aminoacyl-ARN<sub>t</sub> synthétase d'un acide aminé sur un ARN<sub>t</sub>, puis une  $\beta$ -substitution sur l'acide aminé estérifié par une enzyme PLP-dépendante, avec un chalcogène activé.

La structure de la SepCysS a été résolue et est visible au §1.4.3, page 58. Cette enzyme est un dimère, proche phylogéniquement des autres enzymes à substrats soufrés que sont les protéines CSD, NifS et IscS, d'après O'Donoghue *et al.*,<sup>[1.185]</sup> ce qui n'est pas en accord avec l'arbre phylogénique présenté en figure 2.9.<sup>[1.259]</sup>

Même si la position phylogénique de la SepCysS par rapport aux cystéine désulfurases du groupe I et II, ou aux sélénocystéine synthases reste trouble, leurs liens ne peuvent

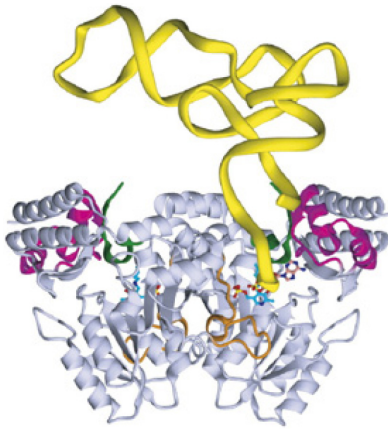


**Figure 2.9** – Arbre phylogénique basé sur des alignements de séquences, contenant des protéines SelA (violet), SepSecS (rouge) et "SelA-like" (vert), parmi d'autres protéines PLP-dépendantes regroupées en famille, dont SepCysS surlignée en jaune.<sup>[1.259]</sup>

qu'être très anciens, proches de la divergence en plusieurs familles des enzymes PLP-dépendantes primordiales et bien avant le stade LUCAS.

Pour la comparaison entre SepCysS et Sela, deux enzymes interagissant avec des  $ARN_t$  différents ne peuvent avoir qu'une filiation très lointaine, à l'image du lien des PylRS et SepRS avec la PheRS primordiale (voir l'arbre phylogénique 1.19, page 55). En effet, les déterminants identitaires des différents  $ARN_t$  sont anciens et conservés, comme cela a été montré pour l' $ARN_t^{Cys}$ ,<sup>[1.172]</sup> mais des enzymes interagissant avec l' $ARN_t$  ont pu évoluer et s'adapter à différents  $ARN_t$  préexistants.

Ainsi, sur le modèle de la filiation entre les aminoacyl- $ARN_t$ -synthétases précédentes, un scénario hypothétique propose que les voies de  $\beta$ -remplacement sur le Sep- $ARN_t^{Cys}$  et



**Figure 2.10** – Modèle d'interaction entre SepCysS et l' $ARN_t^{Cys}$ .<sup>[1.174]</sup>

sur le Sep- $ARN_t^{Sec}$ , ou Ser- $ARN_t^{Sec}$ , pourraient être issue de la même voie ancestrale, impliquant une enzyme PLP-dépendante primordiale ayant ensuite évolué et divergé en SepCysS et SepSecS/Sela. Dans cet esprit, il est intéressant de noter que le modèle d'interaction entre la SepCysS et le Sep- $ARN_t^{Cys}$  proposé par R. Fukunaga *et al.*<sup>[1.174]</sup> fait intervenir une reconnaissance du bras accepteur de l' $ARN_t$  (modèle reproduit figure 2.10). Or le caractère le plus discriminant dans l'interaction entre Sela et le séryl- $ARN_t^{Sec}$  semble également être le bras accepteur de ce dernier (voir les études des déterminants de l' $ARN_t^{Sec}$  au § 2.1.1, page 83).

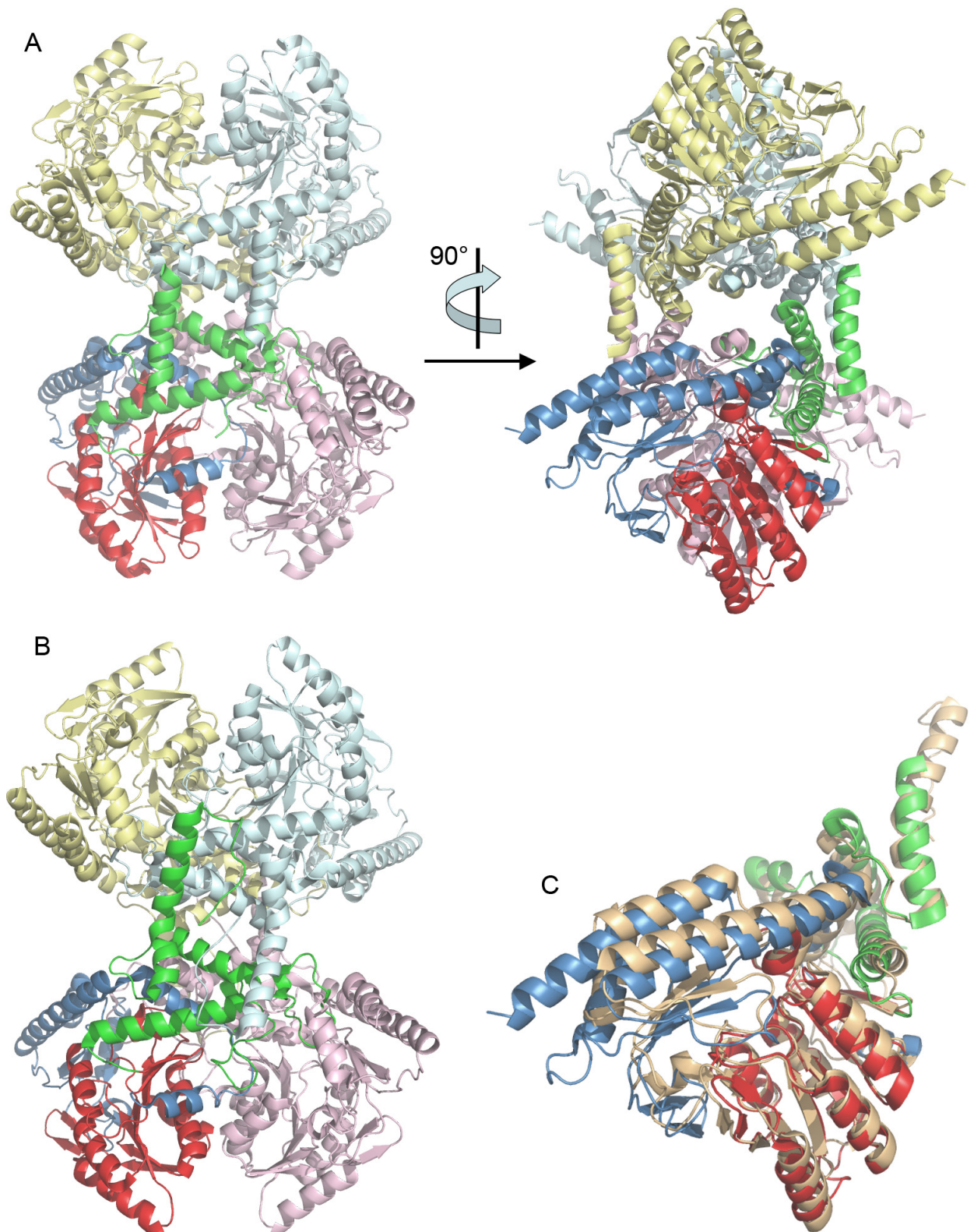
Pourtant, aucun élément ne permet de relier les  $ARN_t^{Sec}$  et  $ARN_t^{Cys}$ , dont les discriminants actuels sont très différents. C'est pourquoi la comparaison des sélénocystéine synthases bactériennes et eucaryotes entre elles paraît encore plus pertinente.

### 2.3.5.2 Le lien logique avec les SepSecS

Si les enzymes SepCysS et Sela/SepSecS peuvent avoir un lien datant de l'émergence du code génétique, les enzymes Sela et SepSecS ont sans doute un ancêtre commun datant de l'époque LUCAS, juste avant la divergence entre le domaine des eubactéries et les deux autres domaines du vivant. En effet, ces deux enzymes catalysent une réaction similaire, conduisant au sélénocystéinyl- $ARN_t^{Sec}$ , à partir de séryl- $ARN_t^{Sec}$  pour Sela, et de phosphoséryl- $ARN_t^{Sec}$  pour SepSecS. Les différents points communs entre les voies de synthèse et d'incorporation de la sélénocystéine chez les procaryotes et eucaryotes montrent qu'elles dérivent d'une voie unique utilisée avant la divergence des domaines.

L'enzyme Sela et le couple kinase PSTK / SepSecS sont d'ailleurs inter-compatibles : les séryl- $ARN_t^{Sec}$  bactériens et eucaryotes réagissent de façon similaire avec l'une ou l'autre voie enzymatique.<sup>[1.228]</sup>

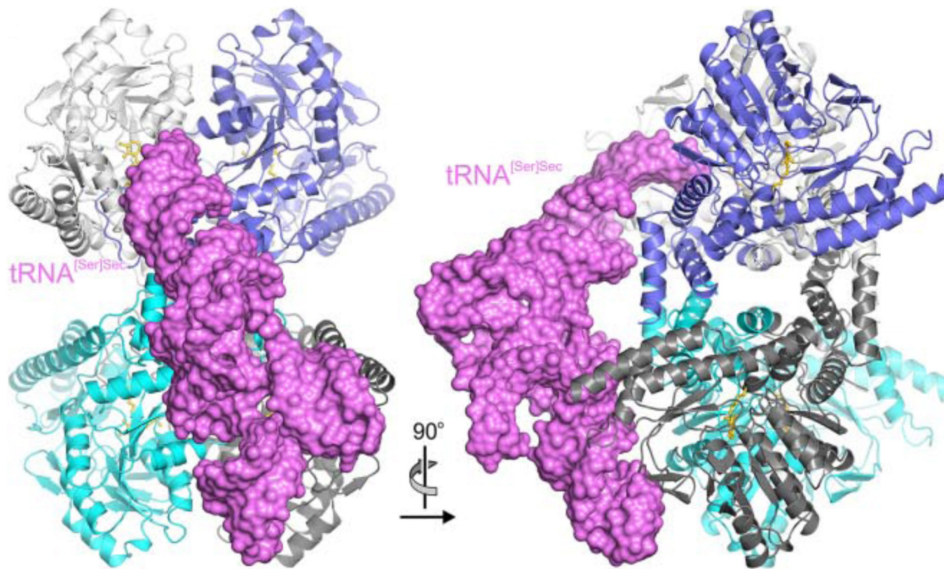
Pourtant, à côté de ce lien apparent, les deux enzymes ne partagent pas la même structure oligomérique : Sela est un décamère et SepSec est un tétramère chez les eucaryotes et chez les archaées, comme le montrent les structures récemment résolues des enzymes de *Mus musculus*<sup>[1.266]</sup> et de *Methanococcus maripaludis*,<sup>[1.260]</sup> présentées figure 2.11.



**Figure 2.11** – Structures tridimensionnelles des SepSecS eucaryote et archaie résolues en 2008.  
 (A) Structure de la SepSecS tétramérique de *Mus musculus*<sup>[1.266]</sup> (n° PDB : 3BC8).  
 (B) Structure de la SepSecS tétramérique de *Methanococcus maripaludis*<sup>[1.260]</sup> (n° PDB : 2Z67).  
 (C) Alignement d'un monomère de SepSecS archaie (orange) avec un monomère de SepSecS eucaryote.  
 (code couleur : vert=N-ter, rouge=Coeur, bleu=C-ter)

Ces structures ont été résolues et publiées en 2008, deux ans après la découverte de l'identité des SepSecS chez les eucaryotes et les archaées.<sup>[I.242,I.259]</sup> Elle montrent d'ailleurs ce que les comparaisons de séquences présageaient : les SepSecS eucaryote et archaées sont très similaires structurellement (voir l'alignement de deux monomères en (C) de la figure 2.11, page précédente).

La structure de la SepSecS eucaryote a même permis à O. Ganichkin *et al.*<sup>[I.266]</sup> de proposer un modèle d'interaction entre le Sep- $ARN_t^{Sec}$  et cette enzyme, reproduit figure 2.12 ci-dessous.

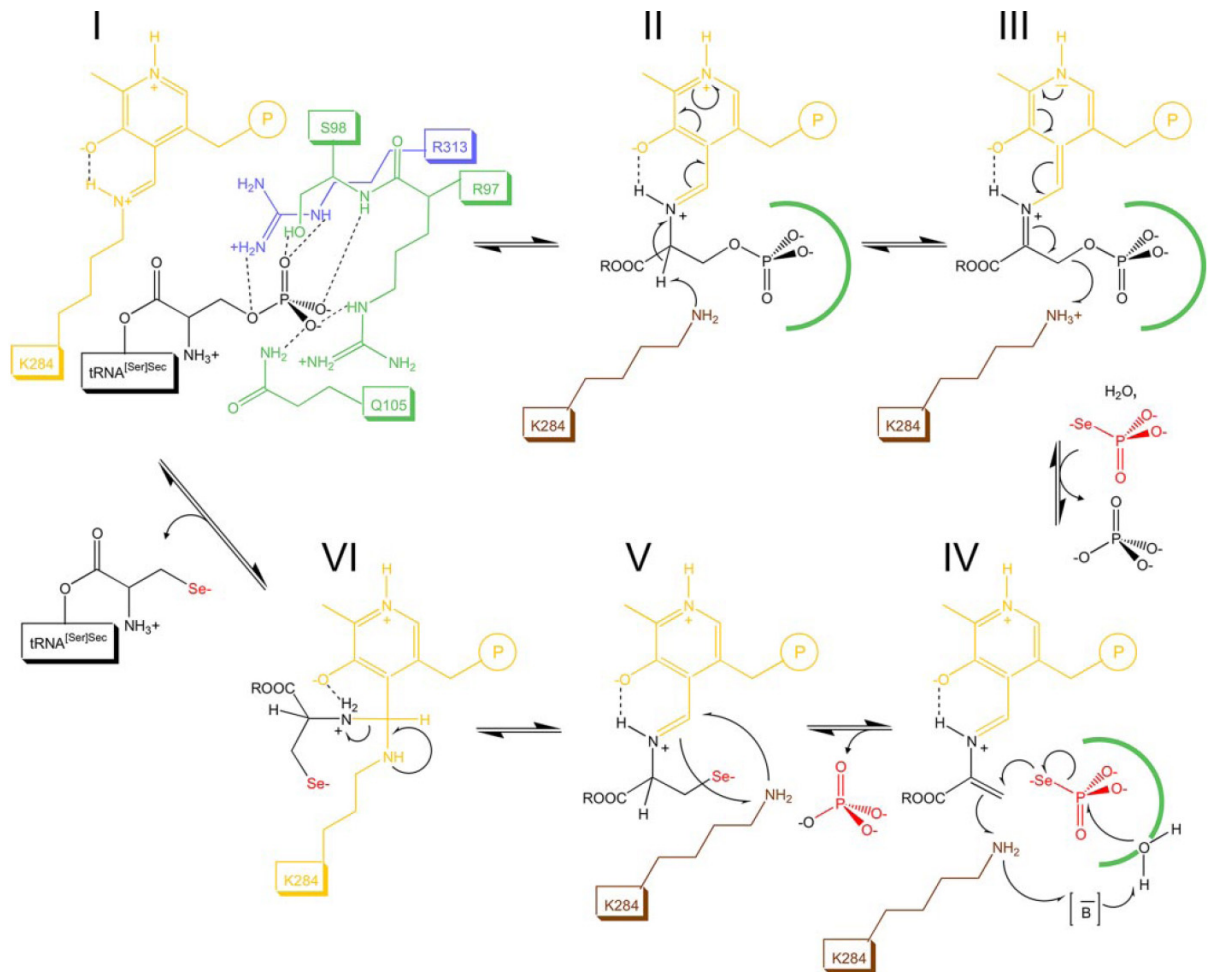


**Figure 2.12** – Modèle d'interaction entre SepSecS et l' $ARN_t^{Sec}$ .<sup>[I.266]</sup>

Dans ce modèle, le positionnement particulier de l' $ARN_t^{Sec}$  tente d'expliquer l'organisation oligomérique particulière de SepSecS, en s'appuyant sur la distance appropriée qui sépare le site actif et les surfaces chargées positivement de la protéine, qui pourrait avoir une forte affinité pour les phosphates du squelette de l' $ARN_t^{Sec}$ , chargés négativement. Ce modèle propose donc que SepSecS reconnaisse l' $ARN_t^{Sec}$  en plusieurs points et par l'intermédiaire de différents protomères du tétramère. Ses auteurs suggèrent même que l'organisation décamerique spécifique de Sela pourrait avoir émergé sur un même principe.

La structuration du site actif de SepSecS permet également aux auteurs de proposer un mécanisme catalytique, présent figure 2.13, page suivante, qui confirme par homologie celui établi pour Sela (retranscrit figure 2.7, page 95).

Une autre information importante a pu être déduite des structures résolues de ces SepSecS. Elles possèdent, entre autres, deux caractéristiques qui les différencient des autres membres de la famille des enzymes PLP-dépendantes de «Fold-Type I» : une extension N-terminale permettant l'organisation oligomérique particulière de l'enzyme, et une large insertion dans sa séquence qui remodèle le site actif.<sup>[I.266]</sup>



**Figure 2.13** – Mécanisme du  $\beta$ -remplacement effectué par SepSecS (Tiré de O. Ganichkin *et al.*<sup>[I.266]</sup>).

Ces caractéristiques particulières font des SepSecS des enzymes à la marge de la famille des enzymes PLP-dépendantes de «Fold-Type I». Elles constituent une nouvelle sous-classe de cette famille. Cette séparation est d'ailleurs visible sur l'arbre phylogénique présent figure 2.9, page 99,<sup>[I.259]</sup> où les SepSecS sont même placées à part par rapport aux enzymes bactériennes Sela.

Il est donc possible que les enzymes SepSecS et Sela, malgré la catalyse d'une même réaction de  $\beta$ -remplacement sur un substrat similaire, aient fortement divergé l'une de l'autre à partir de leur possible ascendant primordial, depuis la séparation entre le domaine des eubactéries et les deux autres domaines du vivant. Cette possibilité implique que les informations structurales obtenues sur les SepSecS ne peuvent être transposées à Sela par homologie.



Une autre entité paraît encore plus proche des SelA bactériennes que ne le sont les SepSecS lors des alignements et comparaisons de séquences, et pourraient nous apporter des indices intéressants sur notre objet d'étude. Il s'agit des protéines archaïques au nom évocateur de "SelA-like".

### 2.3.5.3 La similarité des "SelA-like"

Les protéines "SelA-like" ont été découvertes lors des recherches sur les potentielles sélénocystéine synthases des eucaryotes et des archaïques, et l'étude fonctionnelle et structurale de l'une d'entre elles fut publiée en 2005 par J. Kaiser *et al.*<sup>[1.265]</sup>

La protéine "SelA-like" est le produit d'un "cadre de lecture ouvert" (ou "Open Reading Frame" ORF) conservé dans certains génomes d'archaïques méthanogènes :

l'ORF *mj0158* chez *Methanococcus jannaschii*,

l'ORF *mmp0002* chez *Methanococcus maripaludis*,

l'ORF *mk0620* chez *Methanopyrus kandleri*,

l'ORF *mth1914* chez *Methanobacterium thermoautotrophicum*.

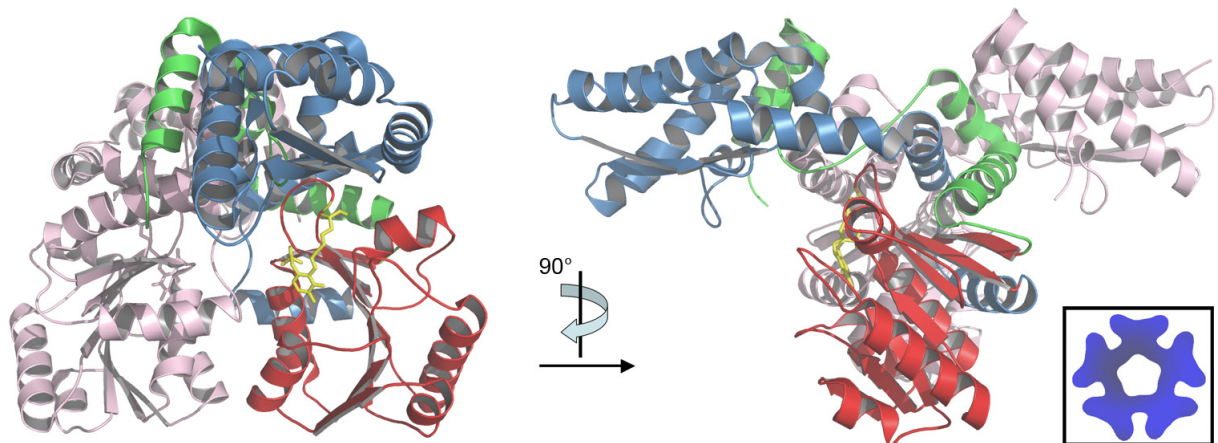
Les trois premiers de ces organismes ont la capacité de synthétiser des sélénoprotéines, alors que le quatrième ne paraît pas avoir cette capacité. Malgré l'exception que représente ce quatrième organisme, la possibilité que les "SelA-like" soient la version archaïque de la sélénocystéine synthase fut envisagée, leurs séquences étant très similaires à celle de SelA, en particulier autour du site actif présumé.

La protéine produite à partir de l'ORF de *M. jannaschii* fut sujette à une étude fonctionnelle et structurale : cette protéine MJ0158 de 374 acides aminés forme un dimère en solution avec une masse apparente en chromatographie sur colonne de gel d'exclusion stérique de 90000 kDa. Il s'agit d'une enzyme PLP-dépendante portant un cofacteur par monomère.

La structure cristallographique de cette protéine fut résolue, comme illustré par la figure 2.14, page ci-contre. Cette structure montre que MJ0158 possède deux domaines : le domaine de coeur en sandwich  $\alpha/\beta/\alpha$  des enzymes PLP-dépendantes et un petit domaine composé des parties N-terminales et C-terminale de la protéine. Cette topologie est similaire aux protéines homologues du domaine NifS, telles que IscS ou CSD. L'arrangement de ces domaines entre eux est par contre relativement différent de ceux rencontrés chez les autres protéines PLP-dépendantes, et permet à MJ0158 d'avoir un site catalytique très ouvert et une forme de T unique.

C'est d'ailleurs cette forme en T qui donne à cette protéine tout l'intérêt structural d'une comparaison avec SelA. En effet, comme le soulignent les auteurs de la structure tridimensionnelle de MJ0158, cette forme peut faire penser à l'empreinte d'un dimère de SelA sur les clichés de cryomicroscopie électronique de ce décimère.

Comme à l'époque, l'identité de SepSecS chez les eucaryotes et les archaïques n'était pas connue, l'activité sélénocystéine synthase de MJ0158 fut testée. Mais cette protéine ne réagit avec aucun séryl-ARN<sub>t</sub><sup>Sec</sup> ou phosphoséryl-ARN<sub>t</sub><sup>Sec</sup>, même en présence de sélénophosphate. Cette protéine ne présente également aucune réactivité avec les 20 acides-



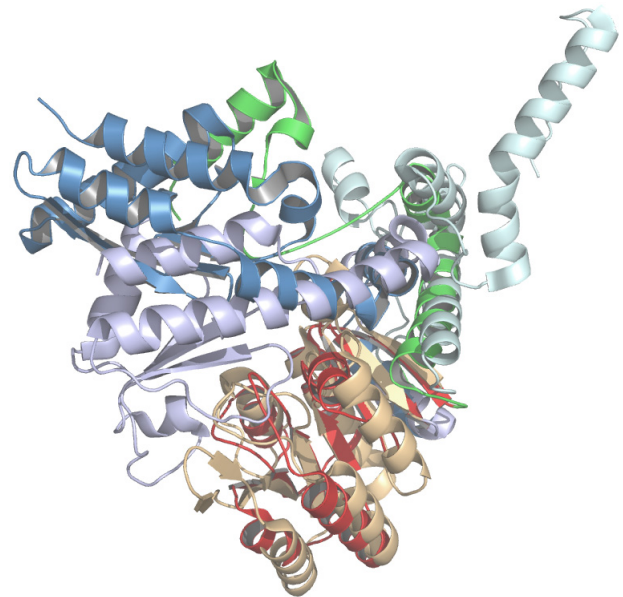
**Figure 2.14** – Structure tridimensionnelle de la “SelA-like” MJ0158 de *M. jannaschii*.<sup>[1.265]</sup> (n° PDB : 2AEU) En encart, l’empreinte de SelA en microscopie électronique.<sup>[1.254]</sup> (code couleur : vert=N-ter, rouge=Coeur, bleu=C-ter)

aminés canoniques libres, ni avec des composés soufrés habituels. La fonction des enzymes “SelA-like” archaïques est donc actuellement inconnue.

Quel est alors l’intérêt d’une comparaison entre ces protéines à l’activité inconnue et SelA ?

Comme indiqué précédemment, l’arrangement structural des domaines de MJ0158 entre eux est relativement particulier.

L’alignement des structures d’un monomère de MJ0158 et de SepSecS archaïques, en figure 2.15, le montre particulièrement bien. Les coeurs  $\alpha/\beta/\alpha$  des deux enzymes ne sont pas très différents l’un de l’autre, SepSecS ayant simplement une séquence plus longue au niveau de ce domaine, ce qui se traduit par des boucles plus importantes et des hélices  $\alpha$  plus étendues. Par contre, les parties N- et C-terminales s’organisent différemment au-dessus de ce coeur conservé : l’hélice  $\alpha$  N-terminale se positionne avec la partie C-terminale, loin de son homologue qui sert chez SepSecS à permettre la formation de tétramères. Il en résulte un petit domaine ayant un angle plus ouvert par rapport au domaine coeur et qui donne à la protéine cette forme de T caractéristique.



**Figure 2.15** – Comparaison des monomères de MJ0158 et de SepSecS archaïques (couleurs claires), par alignement du coeur  $\alpha/\beta/\alpha$ .

Bien que la séquence de MJ0158 soit également plus courte que celle de SelA, il s’agit de la séquence protéique ayant le plus d’identité avec SelA, dans l’état actuel des

connaissances. Il se pourrait donc que cette protéine d'activité inconnue soit actuellement le meilleur aperçu de ce à quoi peut ressembler SelA, en version dimérique.

### 2.3.6 Rares sont les structures décāmériques

SelA a une organisation structurale très particulière, la rendant unique dans la famille des enzymes PLP-dépendantes et ne permettant pas de trouver de modèle valide de sa structure.

De manière générale, les enzymes homo-décāmériques possédant une symétrie rotationnelle d'ordre 5 sont rares. Depuis les deux dernières décennies, des données sont disponibles sur la structure de certaines de ces protéines marginales, résumées dans le tableau 2.5.

Protéine	Organisme	Masse du monomère	Identification PDB
Muconolactone isomérase	<i>Pseudomonas putida</i>	11 kDa	1MLI <sup>[I.267]</sup>
Méthanol-déshydrogénase	<i>Bacillus subtilis</i>	43 kDa	_ <sup>[I.268]</sup>
“Bovine pancreatic trypsin inhibitor”	<i>Bos taurus</i>	6,5 kDa	1BHC <sup>[I.269]</sup>
Cyanase	<i>Escherichia coli</i>	17 kDa	1DW9 <sup>[I.270]</sup>
2-Cys peroxyrédoxine	<i>Homo Sapiens</i>	22 kDa	1QMV <sup>[I.271]</sup>
Fructose-6-phosphate aldolase	<i>Escherichia coli</i>	25 kDa	1L6W <sup>[I.272]</sup>
GTP cyclohydrolase I	<i>Rattus norvegicus</i>	27 kDa	1HS7 <sup>[I.273]</sup>
Ketopantoate hydroxyméthyltransférase	<i>M. tuberculosis</i>	29 kDa	1OY0 <sup>[I.274]</sup>
Lectine galactose-spécifique	<i>Crotalus atrox</i>	15 kDa	1JZN <sup>[I.275]</sup>
Lumazine synthase (RibH2)	<i>Brucella abortus</i>	18 kDa	1XN1 <sup>[I.276]</sup>

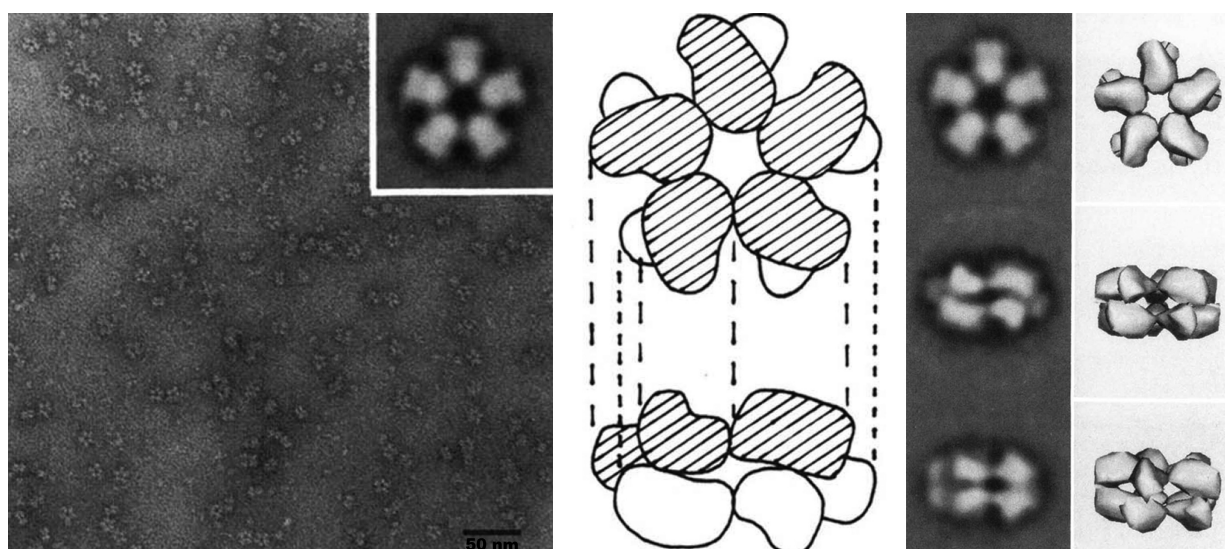
**Tableau 2.5** – Protéines homo-décāmériques possédant une symétrie rotationnelle d'ordre 5.

Sauf dans le cas des peroxyrédoxines, ces enzymes homo-décāmériques sont des protéines exotiques, n'ayant qu'un seul représentant possédant cette structure et ne dépassant pas 30 kDa. Les seules exceptions sont SelA avec 50 kDa et la Méthanol-déshydrogénase de *Bacillus subtilis* avec 43 kDa, deux protéines dont les structures ne sont pas encore résolues, mais qui ont été étudiées en cryomicroscopie électronique (voir les clichés obtenus pour cette dernière en 2.16, page suivante).

Dans le cas de la grande famille des peroxyrédoxines, les peroxyrédoxines de type 2-Cys existent sous forme de dimères et de décāmères dans la cellule, et son état d'oligomérisation est fonction de son état d'oxydo-réduction.<sup>[I.277]</sup> Il s'agit là d'un cas particulier d'homo-décāmère répandu.

Les structures résolues de ces enzymes marginales sont visibles en Annexe A.7, page 241. Mais comme aucune de ces protéines n'appartient à la famille des enzymes PLP-dépendantes, leur structures n'apportent pas plus d'indice sur l'organisation de SelA.

SelA paraît être la seule de ces enzymes à arborer une organisation décāmérique et appartient donc bien à une sous-classe particulière de la famille des enzymes PLP-dépendantes de «Fold-Type I».



**Figure 2.16** – Cliché de cryo-microscopie électronique et projections de la Méthanol-déshydrogénase de *Bacillus subtilis*.<sup>[1.268]</sup>

## 2.4 Conclusion

Nous venons de passer en revue les connaissances actuelles sur l'exception au code génétique qu'est la sélénocystéine. Cet acide aminé est à la fois particulier par son mode de biosynthèse et par son mode d'incorporation lors de la traduction protéique.

Les trois domaines du vivant ont cette exception en commun. Le système complexe mis en place pour arriver à synthétiser des sélénoprotéines, aujourd'hui différent chez les eubactéries et chez les eucaryotes et les archaées, semble bien être issu d'un même mécanisme primordial, apparu entre l'émergence du code génétique et la séparation des domaines du vivant, post-LUCAS. La forte dynamique des phénomènes de perte et de gain de la capacité de synthétiser la sélénocystéine est intimement liée aux conditions environnementales subies par les organismes, et aux pressions génétiques qui en découlent.

Les questions évolutives sous-jacentes au cas de la sélénocystéine sont liées :

- au positionnement de l'émergence de l'utilisation de cet acide aminé par rapport aux autres acides aminés du code génétique, en particulier par rapport à l'ajout tardif de la glutamine, de l'asparagine et de la cystéine.

- à l'utilisation du codon UGA pour le recodage de la sélénocystéine, en particulier par rapport à sa signification ambiguë dans la plupart des génomes. Quel est le premier sens à avoir été donné à ce codon, et de quelle manière a émergé le système de recodage cotraductionnel, permettant une levée d'ambiguïté message-dépendant ?

Les réponses à ces questions ne peuvent être obtenues directement pas une étude menée sur les acteurs de la biosynthèse de la sélénocystéine, mais celle-ci peut permettre d'apporter des indices sur le passé de l'utilisation de cet acide aminé.

Ainsi l'étude de la sélénocystéine synthase bactérienne peut avoir pour ligne de mire l'éclaircissement des liens entre cette protéine et les autres membres de la famille des enzymes PLP-dépendantes de «Fold-Type I», en particulier avec les SepCysS et “Sela-like” archaees, mais aussi avec ses cousines SepSecS eucaryote et archaee.

Cette étude peut aussi essayer de répondre aux questions particulières qui ont pour sujet :

- la structure décamerique de Sela. Pourquoi cette structure spécifique et exotique pour une enzyme PLP-dépendante ? A-t-elle une importance fonctionnelle ?

- l'interaction entre l' $ARN_t^{Sec}$  et Sela. Quelles sont les zones d'interactions entre la protéine et l'ARN ? Quels est le sens de la distribution des  $ARN_t^{Sec}$  sur un décamerè protéique ? Quelles sont les changements conformationnels que peut induire l' $ARN_t^{Sec}$  lorsqu'il interagit avec une sous-unité dimérique de Sela ?

- les données enzymatiques sur Sela. Existe-t-il une coopérativité entre la zone de reconnaissance et le site actif, entre les deux sites actifs d'un dimère, entre les différents dimères du décamerè ?

La technique de choix pour obtenir le plus d'information sur le couple formé par l' $ARN_t^{Sec}$  et Sela est la cristallographie structurale, qui permettrait d'accéder à une vue d'ensemble des éléments de la biosynthèse de la sélénocystéine et de son fonctionnement.

La caractérisation biochimique du fonctionnement du couple Sela /  $ARN_t^{Sec}$  peut également permettre d'obtenir des informations sur les mécanismes à l'oeuvre lors de cette biosynthèse.

C'est donc dans ces deux voies que les efforts expérimentaux de ce travail de doctorat ont été portés.

# Références bibliographiques de la Partie I

- [1] CRICK F. H. : The origin of the genetic code. *Journal of molecular biology*, 38(3) : p367–379, Dec 1968.
- [2] ZINONI F. ; BIRKMANN A. ; LEINFELDER W. ; BÖCK A. : Cotranslational insertion of selenocysteine into formate dehydrogenase from *Escherichia coli* directed by a UGA codon. *Proceedings of the National Academy of Sciences of the United States of America*, 84(10) : p3156–3160, May 1987.
- [3] LEINFELDER W. ; FORCHHAMMER K. ; ZINONI F. ; SAWERS G. ; MANDRAND-BERTHELOT M. A. ; BÖCK A. : *Escherichia coli* genes whose products are involved in selenium metabolism. *Journal of bacteriology*, 170(2) : p540–546, Feb 1988.
- [4] LEINFELDER W. ; ZEHELEIN E. ; MANDRAND-BERTHELOT M. A. ; BÖCK A. : Gene for a novel tRNA species that accepts L-serine and cotranslationally inserts selenocysteine. *Nature*, 331(6158) : p723–725, Feb 1988.
- [5] DUVOID I. : *Le Sélénium, un élément essentiel parfois redoutable ou le rapport bénéfice/risque du sélénium*. 99 pages. Rapport de recherche bibliographique de l'ENSSIB, 1999.
- [6] Selenium. *Mining Journal Review of the Minor Metals Trade Association MMTA*, Jan 2006.
- [7] STADTMAN T. C. : Some selenium-dependent biochemical processes. *Advances in enzymology and related areas of molecular biology*, 48 : p1–28, 1979.
- [8] TERRY N. ; ZAYED A. M. ; DE SOUZA M. P. ; TARUN A. S. : Selenium in Higher Plants. *Annual Review of Plant Physiology and Plant Molecular Biology*, 51 : p401–432, Jun 2000.
- [9] BIRINGER M. ; PILAWA S. ; FLOHÉ L. : Trends in selenium biochemistry. *Natural product reports*, 19(6) : p693–718, Dec 2002.
- [10] STADTMAN T. C. : Selenium biochemistry. *Annual review of biochemistry*, 59 : p111–127, 1990.
- [11] STADTMAN T. C. : Selenium-dependent enzymes. *Annual review of biochemistry*, 49 : p93–110, 1980.
- [12] PINSENT J. : The need for selenite and molybdate in the formation of formic dehydrogenase by members of the coli-aerogenes group of bacteria. *The Biochemical journal*, 57(1) : p10–16, May 1954.
- [13] PATTERSON E. L. ; MILSTREY R. ; STOKSTAD E. L. : Effect of selenium in preventing exudative diathesis in chicks. *Proceedings of the Society for Experimental Biology and Medicine*, 95(4) : p617–620, Aug 1957.
- [14] SCHWARZ K. ; FOLTZ C. : Selenium as an Integral Part of Factor 3 against Dietary Necrotic Liver Degeneration. *Journal of the American Chemical Society*, 79(12) : p3292–3293, 1957.
- [15] FU L.-H. ; WANG X.-F. ; EYAL Y. ; SHE Y.-M. ; DONALD L. J. ; STANDING K. G. ; BEN-HAYYIM G. : A selenoprotein in the plant kingdom. Mass spectrometry confirms that an opal codon (UGA) encodes selenocysteine in *Chlamydomonas reinhardtii* glutathione peroxidase. *The Journal of biological chemistry*, 277(29) : p25983–25991, Jul 2002.

- [16] SHIGEOKA S.; TAKEDA T. ; HANAOKA T. : Characterization and immunological properties of selenium-containing glutathione peroxidase induced by selenite in *Chlamydomonas reinhardtii*. *The Biochemical journal*, 275(Pt 3) : p623–627, May 1991.
- [17] YOKOTA A.; SHIGEOKA S.; ONISHI T. ; KITAOKA S. : Selenium as Inducer of Glutathione Peroxidase in low-CO(2)-Grown *Chlamydomonas reinhardtii*. *Plant Physiol*, 86(3) : p649–651, Mar 1988.
- [18] HUANG K.; LAURIDSEN E. ; CLAUSEN J. : Selenium-containing peroxidases of germinating barley. *Biological trace element research*, 46(1-2) : p173–182, Oct 1994.
- [19] SABEH F.; WRIGHT T. ; NORTON S. J. : Purification and characterization of a glutathione peroxidase from the Aloe vera plant. *Enzyme & protein*, 47(2) : p92–98, 1993.
- [20] HUANG K. X.; AN Y. X.; CHEN Z. X. ; XU H. B. : Isolation and partial characterization of selenium-containing tRNA from germinating barley. *Biological trace element research*, 82(1-3) : p247–257, 2001.
- [21] FREEMAN J. L.; LINDBLOM S. D.; QUINN C. F.; FAKRA S.; MARCUS M. A. ; PILON-SMITS E. A. H. : Selenium accumulation protects plants from herbivory by Orthoptera via toxicity and deterrence. *The New phytologist*, 175(3) : p490–500, 2007.
- [22] NEUHIERL B.; BÖCK A. : On the mechanism of selenium tolerance in selenium-accumulating plants. Purification and characterization of a specific selenocysteine methyltransferase from cultured cells of *Astragalus bisculatus*. *European journal of biochemistry / FEBS*, 239(1) : p235–238, Jul 1996.
- [23] BURNELL J. N. : Cysteinyl-tRNA Synthetase from *Astragalus* Species. *Plant Physiol*, 63(6) : p1095–1097, Jun 1979.
- [24] FRANKENBERGER J. W. T.; ARSHAD M. : Bioremediation of selenium-contaminated sediments and water. *BioFactors*, 14(1-4) : p241–254, 2001.
- [25] TURNER D. C.; STADTMAN T. C. : Purification of protein components of the clostridial glycine reductase system and characterization of protein A as a selenoprotein. *Archives of biochemistry and biophysics*, 154(1) : p366–381, Jan 1973.
- [26] ANDREESEN J. R.; LJUNGDAHL L. G. : Formate dehydrogenase of *Clostridium thermoaceticum* : incorporation of selenium-75, and the effects of selenite, molybdate, and tungstate on the enzyme. *Journal of bacteriology*, 116(2) : p867–873, Nov 1973.
- [27] ROTRUCK J. T.; POPE A. L.; GANTHER H. E.; SWANSON A. B.; HAFEMAN D. G. ; HOEKSTRA W. G. : Selenium : biochemical role as a component of glutathione peroxidase. *Science*, 179(73) : p588–590, Feb 1973.
- [28] FLOHE L.; GÜNZLER W. A. ; SCHOCK H. H. : Glutathione peroxidase : a selenoenzyme. *FEBS letters*, 32(1) : p132–134, May 1973.
- [29] MILLS G. C.; RANDALL H. P. : Hemoglobin catabolism. II. The protection of hemoglobin from oxidative breakdown in the intact erythrocyte. *The Journal of biological chemistry*, 232(2) : p589–598, Jun 1958.
- [30] O'BRIEN P. J.; LITTLE C. : Intracellular mechanisms for the decomposition of a lipid peroxide. II. Decomposition of a lipid peroxide by subcellular fractions. *Canadian journal of biochemistry*, 47(5) : p493–499, May 1969.
- [31] SCHNEIDER F.; FLOHÉ L. : Studies on glutathione : H<sub>2</sub>O<sub>2</sub>-oxidoreductase (glutathione peroxidase). *Hoppe-Seyler's Zeitschrift für physiologische Chemie*, 348(5) : p540–552, May 1967.
- [32] GÜNZLER W. A.; VERGIN H.; MÜLLER I. ; FLOHÉ L. : Glutathione peroxidase VI : the reaction of glutathione peroxidase with various hydroperoxides. *Hoppe-Seyler's Zeitschrift für physiologische Chemie*, 353(6) : p1001–1004, Jun 1972.
- [33] SHAMBERGER R. J.; FROST D. V. : Possible protective effect of selenium against human cancer. *Canadian Medical Association journal*, 100(14) : p682, Apr 1969.
- [34] SHAMBERGER R. J.; WILLIS C. E. : Selenium distribution and human cancer mortality. *CRC critical reviews in clinical laboratory sciences*, 2(2) : p211–221, Jun 1971.

- 
- [35] COMBS J. G. F. ; COMBS S. B. : Selenium effects on drug and foreign compound toxicity. *Pharmacology & therapeutics*, 33(2-3) : p303–315, 1987.
- [36] COMBS J. G. F. ; GRAY W. P. : Chemopreventive agents : selenium. *Pharmacology & therapeutics*, 79(3) : p179–192, Sep 1998.
- [37] COMBS J. G. F. ; CLARK L. C. ; TURNBULL B. W. : An analysis of cancer prevention by selenium. *BioFactors*, 14(1-4) : p153–159, 2001.
- [38] GANTHER H. E. : Selenium metabolism, selenoproteins and mechanisms of cancer prevention : complexities with thioredoxin reductase. *Carcinogenesis*, 20(9) : p1657–1666, Sep 1999.
- [39] YU S. Y. ; ZHU Y. J. ; LI W. G. ; HUANG Q. S. ; HUANG C. Z. ; ZHANG Q. N. ; HOU C. : A preliminary report on the intervention trials of primary liver cancer in high-risk populations with nutritional supplementation of selenium in China. *Biological trace element research*, 29(3) : p289–294, Jun 1991.
- [40] CLARK L. C. ; COMBS J. G. F. ; TURNBULL B. W. ; SLATE E. H. ; CHALKER D. K. ; CHOW J. ; DAVIS L. S. ; GLOVER R. A. ; GRAHAM G. F. ; GROSS E. G. ; KRONGRAD A. ; LESHER J. J. L. ; PARK H. K. ; SANDERS J. B. B. ; SMITH C. L. ; TAYLOR J. R. : Effects of selenium supplementation for cancer prevention in patients with carcinoma of the skin. A randomized controlled trial. Nutritional Prevention of Cancer Study Group. *JAMA : the journal of the American Medical Association*, 276(24) : p1957–1963, Dec 1996.
- [41] HOFFMANN P. R. : Mechanisms by which selenium influences immune responses. *Archivum immunologiae et therapiae experimentalis*, 55(5) : p289–297, Sep 2007.
- [42] SHERIDAN P. A. ; ZHONG N. ; CARLSON B. A. ; PERELLA C. M. ; HATFIELD D. L. ; BECK M. A. : Decreased selenoprotein expression alters the immune response during influenza virus infection in mice. *The Journal of nutrition*, 137(6) : p1466–1471, Jun 2007.
- [43] STEHBENS W. E. : Oxidative stress in viral hepatitis and AIDS. *Experimental and molecular pathology*, 77(2) : p121–132, Oct 2004.
- [44] HURWITZ B. E. ; KLAUS J. R. ; LLABRE M. M. ; GONZALEZ A. ; LAWRENCE P. J. ; MAHER K. J. ; GREESON J. M. ; BAUM M. K. ; SHOR-POSNER G. ; SKYLER J. S. ; SCHNEIDERMAN N. : Suppression of human immunodeficiency virus type 1 viral load with selenium supplementation : a randomized controlled trial. *Archives of Internal Medicine*, 167(2) : p148–154, Jan 2007.
- [45] PASSARETTI C. ; GUPTA A. : Selenium and HIV-1 : hope or hype? *Archives of internal medicine*, 167(22) : p2530–2531, Dec 2007.
- [46] ROSS D. A. ; COUSENS S. ; WEDNER S. H. ; SISMANIDIS C. : Does selenium supplementation slow progression of HIV ? : Potentially misleading presentation of the results of a trial. *Archives of internal medicine*, 167(14) : p1555–6; author reply 1557, Jul 2007.
- [47] DILLON D. M. B. ; STAPLETON J. T. : Selenium effects on HIV RNA and CD4 cell counts. *Archives of internal medicine*, 167(14) : p1556–7; author reply 1557, Jul 2007.
- [48] GLADYSHEV V. N. ; STADTMAN T. C. ; HATFIELD D. L. ; JEANG K. T. : Levels of major selenoproteins in T cells decrease during HIV infection and low molecular mass selenium compounds increase. *Proceedings of the National Academy of Sciences of the United States of America*, 96(3) : p835–839, Feb 1999.
- [49] DUNTAS L. H. : The role of selenium in thyroid autoimmunity and cancer. *Thyroid : official journal of the American Thyroid Association*, 16(5) : p455–460, May 2006.
- [50] GÄRTNER R. ; GASNIER B. C. H. : Selenium in the treatment of autoimmune thyroiditis. *BioFactors*, 19(3-4) : p165–170, 2003.
- [51] MAZOKOPAKIS E. E. ; PAPADAKIS J. A. ; PAPADOMANOLAKI M. G. ; BATISTAKIS A. G. ; GIANNAKOPOULOS T. G. ; PROTOPAPADAKIS E. E. ; GANOTAKIS E. S. : Effects of 12 months treatment with L-selenomethionine on serum anti-TPO Levels in Patients with Hashimoto’s thyroiditis. *Thyroid : official journal of the American Thyroid Association*, 17(7) : p609–612, Jul 2007.



- [52] KARANIKAS G. ; SCHUETZ M. ; KONTUR S. ; DUAN H. ; KOMMATA S. ; SCHOEN R. ; ANTONI A. ; KLETTER K. ; DUDCZAK R. ; WILHEIM M. : No immunological benefit of selenium in consecutive patients with autoimmune thyroiditis. *Thyroid : official journal of the American Thyroid Association*, 18(1) : p7–12, Jan 2008.
- [53] GÄRTNER R. ; ALBRICH W. ; ANGSTWURM M. W. : The effect of a selenium supplementation on the outcome of patients with severe systemic inflammation, burn and trauma. *BioFactors*, 14(1-4) : p199–204, 2001.
- [54] ANGSTWURM M. W. A. ; ENGELMANN L. ; ZIMMERMANN T. ; LEHMANN C. ; SPES C. H. ; ABEL P. ; STRAUSS R. ; MEIER-HELLMANN A. ; INSEL R. ; RADKE J. ; SCHÜTTLER J. ; GÄRTNER R. : Selenium in Intensive Care (SIC) : results of a prospective randomized, placebo-controlled, multiple-center study in patients with severe systemic inflammatory response syndrome, sepsis, and septic shock. *Critical care medicine*, 35(1) : p118–126, Jan 2007.
- [55] FLOHÉ L. : Selenium in mammalian spermiogenesis. *Biological chemistry*, 388(10) : p987–995, Oct 2007.
- [56] MAIORINO M. ; FLOHÉ L. ; ROVERI A. ; STEINERT P. ; WISSING J. B. ; URSINI F. : Selenium and reproduction. *BioFactors*, 10(2-3) : p251–256, 1999.
- [57] OLSON G. E. ; WINFREY V. P. ; NAGDAS S. K. ; HILL K. E. ; BURK R. F. : Selenoprotein P is required for mouse sperm development. *Biology of reproduction*, 73(1) : p201–211, Jul 2005.
- [58] REDERSTORFF M. ; KROL A. ; LESCURE A. : Understanding the importance of selenium and selenoproteins in muscle function. *Cellular and molecular life sciences : CMLS*, 63(1) : p52–59, Jan 2006.
- [59] REDERSTORFF M. ; ALLAMAND V. ; GUICHENEY P. ; GARTIOUX C. ; RICHARD P. ; CHAIGNE D. ; KROL A. ; LESCURE A. : Ex vivo correction of selenoprotein N deficiency in rigid spine muscular dystrophy caused by a mutation in the selenocysteine codon. *Nucleic acids research*, 36(1) : p237–244, Jan 2008.
- [60] KRYUKOV G. V. ; GLADYSHEV V. N. : The prokaryotic selenoproteome. *EMBO reports*, 5(5) : p538–543, May 2004.
- [61] COPELAND P. R. : Making sense of nonsense : the evolution of selenocysteine usage in proteins. *Genome biology*, 6(6) : 221, 2005.
- [62] ZHANG Y. ; FOMENKO D. E. ; GLADYSHEV V. N. : The microbial selenoproteome of the Sargasso Sea. *Genome biology*, 6(4) : R37, 2005.
- [63] KRYUKOV G. V. ; KUMAR R. A. ; KOC A. ; SUN Z. ; GLADYSHEV V. N. : Selenoprotein R is a zinc-containing stereo-specific methionine sulfoxide reductase. *Proc Natl Acad Sci U S A*, 99(7) : p4245–4250, Apr 2002.
- [64] KWON S. Y. ; BADENHORST P. ; MARTIN-ROMERO F. J. ; CARLSON B. A. ; PATERSON B. M. ; GLADYSHEV V. N. ; LEE B. J. ; HATFIELD D. L. : The *Drosophila* selenoprotein BthD is required for survival and has a role in salivary gland development. *Molecular and cellular biology*, 23(23) : p8495–8504, Dec 2003.
- [65] CHEN C. L. ; SHIM M. S. ; CHUNG J. ; YOO H.-S. ; HA J. M. ; KIM J. Y. ; CHOI J. ; ZANG S. L. ; HOU X. ; CARLSON B. A. ; HATFIELD D. L. ; LEE B. J. : G-rich, a *Drosophila* selenoprotein, is a Golgi-resident type III membrane protein. *Biochemical and biophysical research communications*, 348(4) : p1296–1301, Oct 2006.
- [66] NOVOSELOV S. V. ; KRYUKOV G. V. ; XU X.-M. ; CARLSON B. A. ; HATFIELD D. L. ; GLADYSHEV V. N. : Selenoprotein H is a nucleolar thioredoxin-like protein with a unique expression pattern. *The Journal of biological chemistry*, 282(16) : p11960–11968, Apr 2007.
- [67] CASTELLANO S. ; GLADYSHEV V. N. ; GUIGÓ R. ; BERRY M. J. : SelenoDB 1.0 : a database of selenoprotein genes, proteins and SECIS elements. *Nucleic acids research*, 36(Database issue) : D332–8, Jan 2008.

- 
- [68] CASTELLANO S. ; LOBANOV A. V. ; CHAPPLE C. ; NOVOSELOV S. V. ; ALBRECHT M. ; HUA D. ; LESCURE A. ; LENGAUER T. ; KROL A. ; GLADYSHEV V. N. ; GUIGÓ R. : Diversity and functional plasticity of eukaryotic selenoproteins : identification and characterization of the SelJ family. *Proceedings of the National Academy of Sciences of the United States of America.*, 102(45) : p16188–16193, Oct 2005.
- [69] LU C. ; QIU F. ; ZHOU H. ; PENG Y. ; HAO W. ; XU J. ; YUAN J. ; WANG S. ; QIANG B. ; XU C. ; PENG X. : Identification and characterization of selenoprotein K : an antioxidant in cardiomyocytes. *FEBS letters*, 580(22) : p5189–5197, Oct 2006.
- [70] SHCHEDRINA V. A. ; NOVOSELOV S. V. ; MALINOUSKI M. Y. ; GLADYSHEV V. N. : Identification and characterization of a selenoprotein family containing a diselenide bond in a redox motif. *Proceedings of the National Academy of Sciences of the United States of America*, 104(35) : p13919–13924, Aug 2007.
- [71] HWANG D. Y. ; SIN J. S. ; KIM M. S. ; YIM S. Y. ; KIM Y. K. ; KIM C. K. ; KIM B. G. ; SHIM S. B. ; JEE S. W. ; LEE S. H. ; BAE C. J. ; LEE B. C. ; JANG M. K. ; CHO J. S. ; CHAE K. R. : Overexpression of human selenoprotein M differentially regulates the concentrations of antioxidants and H<sub>2</sub>O<sub>2</sub>, the activity of antioxidant enzymes, and the composition of white blood cells in a transgenic rat. *International journal of molecular medicine*, 21(2) : p169–179, Feb 2008.
- [72] GRUMOLATO L. ; GHZILI H. ; MONTERO-HADJADJE M. ; GASMAN S. ; LESAGE J. ; TANGUY Y. ; GALAS L. ; AIT-ALI D. ; LEPRINCE J. ; GUÉRINEAU N. C. ; ELKAHLOUN A. G. ; FOURNIER A. ; VIEAU D. ; VAUDRY H. ; ANOUAR Y. : Selenoprotein T is a PACAP-regulated gene involved in intracellular Ca<sup>2+</sup> mobilization and neuroendocrine secretion. *FASEB J*, Jan 2008.
- [73] CASTELLANO S. ; NOVOSELOV S. V. ; KRYUKOV G. V. ; LESCURE A. ; BLANCO E. ; KROL A. ; GLADYSHEV V. N. ; GUIGÓ R. : Reconsidering the evolution of eukaryotic selenoproteins : a novel nonmammalian family with scattered phylogenetic distribution. *EMBO reports*, 5(1) : p71–77, Jan 2004.
- [74] DIKIY A. ; NOVOSELOV S. V. ; FOMENKO D. E. ; SENGUPTA A. ; CARLSON B. A. ; CERNY R. L. ; GINALSKI K. ; GRISHIN N. V. ; HATFIELD D. L. ; GLADYSHEV V. N. : SelT, SelW, SelH, and Rdx12 : genomics and molecular insights into the functions of selenoproteins of a novel thioredoxin-like family. *Biochemistry*, 46(23) : p6871–6882, Jun 2007.
- [75] HATFIELD D. L. ; GLADYSHEV V. N. : How selenium has altered our understanding of the genetic code. *Molecular and cellular biology*, 22(11) : p3565–3576, Jun 2002.
- [76] CONE J. E. ; DEL RÍO R. M. ; DAVIS J. N. ; STADTMAN T. C. : Chemical characterization of the selenoprotein component of clostridial glycine reductase : identification of selenocysteine as the organoselenium moiety. *Proceedings of the National Academy of Sciences of the United States of America*, 73(8) : p2659–2663, Aug 1976.
- [77] LOBANOV A. V. ; FOMENKO D. E. ; ZHANG Y. ; SENGUPTA A. ; HATFIELD D. L. ; GLADYSHEV V. N. : Evolutionary dynamics of eukaryotic selenoproteomes : large selenoproteomes may associate with aquatic life and small with terrestrial life. *Genome biology*, 8(9) : R198, 2007.
- [78] BURK R. F. ; HILL K. E. ; MOTLEY A. K. : Plasma selenium in specific and non-specific forms. *BioFactors*, 14(1-4) : p107–114, 2001.
- [79] HOLLENBACH B. ; MORGENTHALER N. G. ; STRUCK J. ; ALONSO C. ; BERGMANN A. ; KÖHRLE J. ; SCHOMBURG L. : New assay for the measurement of selenoprotein P as a sepsis biomarker from serum. *Journal of trace elements in medicine and biology : organ of the Society for Minerals and Trace Elements (GMS)*, 22(1) : p24–32, 2008.
- [80] STEINBRENNER H. ; ALILI L. ; STUHLMANN D. ; SIES H. ; BRENNEISEN P. : Post-translational processing of selenoprotein P : implications of glycosylation for its utilisation by target cells. *Biological chemistry*, 388(10) : p1043–1051, Oct 2007.
- [81] TRAUlsen H. ; STEINBRENNER H. ; BUCHCZYK D. P. ; KLOTZ L.-O. ; SIES H. : Selenoprotein P protects low-density lipoprotein against oxidation. *Free radical research*, 38(2) : p123–128, Feb 2004.

- [82] BURK R. F. ; HILL K. E. : Selenoprotein P : an extracellular protein with unique physical characteristics and a role in selenium homeostasis. *Annual review of nutrition*, 25 : p215–235, 2005.
- [83] SCHWEIZER U. ; STRECKFUSS F. ; PELT P. ; CARLSON B. A. ; HATFIELD D. L. ; KÖHRLE J. ; SCHOMBURG L. : Hepatically derived selenoprotein P is a key factor for kidney but not for brain selenium supply. *The Biochemical journal*, 386(Pt 2) : p221–226, Mar 2005.
- [84] STEINBRENNER H. ; ALILI L. ; BILGIC E. ; SIES H. ; BRENNISEN P. : Involvement of selenoprotein P in protection of human astrocytes from oxidative damage. *Free radical biology & medicine*, 40(9) : p1513–1523, May 2006.
- [85] STEINBRENNER H. ; BILGIC E. ; ALILI L. ; SIES H. ; BRENNISEN P. : Selenoprotein P protects endothelial cells from oxidative damage by stimulation of glutathione peroxidase expression and activity. *Free radical research*, 40(9) : p936–943, Sep 2006.
- [86] BURK R. F. ; HILL K. E. ; MOTLEY A. K. ; AUSTIN L. M. ; NORSWORTHY B. K. : Deletion of selenoprotein P upregulates urinary selenium excretion and depresses whole-body selenium content. *Biochimica et biophysica acta*, 1760(12) : p1789–1793, Dec 2006.
- [87] PETIT N. ; LESCURE A. ; REDERSTORFF M. ; KROL A. ; MOGHADASZADEH B. ; WEWER U. M. ; GUICHENEY P. : Selenoprotein N : an endoplasmic reticulum glycoprotein with an early developmental expression pattern. *Human molecular genetics*, 12(9) : p1045–1053, May 2003.
- [88] DENIZIAK M. ; THISSE C. ; REDERSTORFF M. ; HINDELANG C. ; THISSE B. ; LESCURE A. : Loss of selenoprotein N function causes disruption of muscle architecture in the zebrafish embryo. *Experimental cell research*, 313(1) : p156–167, Jan 2007.
- [89] GROSJEAN H. ; MARCK C. ; DE CRÉCY-LAGARD V. : The various strategies of codon decoding in organisms of the three domains of life : evolutionary implications. *Nucleic acids symposium series*, (51) : p15–16, 2007.
- [90] WITTWER A. J. ; CHING W. M. : Selenium-containing tRNA(Glu) and tRNA(Lys) from *Escherichia coli* : purification, codon specificity and translational activity. *BioFactors*, 2(1) : p27–34, Mar 1989.
- [91] ROMERO H. ; ZHANG Y. ; GLADYSHEV V. N. ; SALINAS G. : Evolution of selenium utilization traits. *Genome biology*, 6(8) : R66, 2005.
- [92] WITTWER A. J. ; TSAI L. ; CHING W. M. ; STADTMAN T. C. : Identification and synthesis of a naturally occurring selenonucleoside in bacterial tRNAs : 5-[(methylamino)methyl]-2-selenouridine. *Biochemistry*, 23(20) : p4650–4655, Sep 1984.
- [93] CHING W. M. ; WITTWER A. J. ; TSAI L. ; STADTMAN T. C. : Distribution of two selenonucleosides among the selenium-containing tRNAs from *Methanococcus vannielii*. *Proceedings of the National Academy of Sciences of the United States of America*, 81(1) : p57–60, Jan 1984.
- [94] MIZUTANI T. ; WATANABE T. ; KANAYA K. ; NAKAGAWA Y. ; FUJIWARA T. : Trace 5-methylaminomethyl-2-selenouridine in bovine tRNA and the selenouridine synthase activity in bovine liver. *Molecular biology reports*, 26(3) : p167–172, Aug 1999.
- [95] VERES Z. ; TSAI L. ; SCHOLZ T. D. ; POLITINO M. ; BALABAN R. S. ; STADTMAN T. C. : Synthesis of 5-methylaminomethyl-2-selenouridine in tRNAs : <sup>31</sup>P NMR studies show the labile selenium donor synthesized by the selD gene product contains selenium bonded to phosphorus. *Proceedings of the National Academy of Sciences of the United States of America*, 89(7) : p2975–2979, Apr 1992.
- [96] EHRENREICH A. ; FORCHHAMMER K. ; TORMAY P. ; VEPREK B. ; BÖCK A. : Selenoprotein synthesis in *E. coli*. Purification and characterisation of the enzyme catalysing selenium activation. *European journal of biochemistry / FEBS*, 206(3) : p767–773, Jun 1992.
- [97] VERES Z. ; STADTMAN T. C. : A purified selenophosphate-dependent enzyme from *Salmonella typhimurium* catalyzes the replacement of sulfur in 2-thiouridine residues in tRNAs with selenium. *Proceedings of the National Academy of Sciences of the United States of America*, 91(17) : p8092–8096, Aug 1994.

- 
- [98] WOLFE M. D. ; AHMED F. ; LACOURCIERE G. M. ; LAUHON C. T. ; STADTMAN T. C. ; LARSON T. J. : Functional diversity of the rhodanese homology domain : the *Escherichia coli* ybbB gene encodes a selenophosphate-dependent tRNA 2-selenouridine synthase. *The Journal of biological chemistry*, 279(3) : p1801–1809, Jan 2004.
- [99] CATON-WILLIAMS J. ; HUANG Z. : Biochemistry of Selenium-Derivatized Naturally Occurring and Unnatural Nucleic Acids. *Chemistry & biodiversity*, 5(3) : p396–407, Mar 2008.
- [100] HOFFMAN J. L. ; MCCONNELL K. P. : The presence of 4-selenouridine in *Escherichia coli* tRNA. *Biochimica et biophysica acta*, 366(1) : p109–113, Sep 1974.
- [101] LOBANOV A. V. ; HATFIELD D. L. ; GLADYSHEV V. N. : Selenoproteinless animals : Selenophosphate synthetase SPS1 functions in a pathway unrelated to selenocysteine biosynthesis. *Protein science*, 17(1) : p176–182, Jan 2008.
- [102] HAFT D. ; SELF W. : Orphan SelD proteins and selenium-dependent molybdenum hydroxylases. *Biology direct*, 3(1) : 4, Feb 2008.
- [103] IUPAC-IUBMB Joint Commission on Biochemical Nomenclature (JCBN) and Nomenclature Committee of IUBMB (NC-IUBMB), newsletter 1999. *European journal of biochemistry / FEBS*, 264(2) : p607–609, Sep 1999.
- [104] BÖCK A. : Selenocysteine. *Encyclopedia of Life Sciences*, 1–6 pages, Nov 2001.
- [105] MCCONNELL K. P. ; VAN LOON E. J. : Distribution of Se75 in serum proteins as determined by paper electrophoresis. *The Journal of biological chemistry*, 212(2) : p747–750, Feb 1955.
- [106] MCCONNELL K. P. ; WABNITZ C. H. : Studies on the fixation of radioselenium in proteins. *The Journal of biological chemistry*, 226(2) : p765–776, Jun 1957.
- [107] SHRIFT A. ; VIRUPAKSHA T. K. : Biosynthesis of Se-methyl-selenocysteine from selenite in selenium-accumulating plants. *Biochimica et biophysica acta*, 71 : p483–485, May 1963.
- [108] STADTMAN T. C. : Selenium biochemistry. *Science*, 183(128) : p915–922, Mar 1974.
- [109] GÜNZLER W. A. ; STEFFENS G. J. ; GROSSMANN A. ; KIM S. M. ; OTTING F. ; WENDEL A. ; FLOHÉ L. : The amino-acid sequence of bovine glutathione peroxidase. *Hoppe-Seyler's Zeitschrift für physiologische Chemie*, 365(2) : p195–212, Feb 1984.
- [110] CHAMBERS I. ; FRAMPTON J. ; GOLDFARB P. ; AFFARA N. ; MCBAIN W. ; HARRISON P. R. : The structure of the mouse glutathione peroxidase gene : the selenocysteine in the active site is encoded by the 'termination' codon, TGA. *The EMBO journal*, 5(6) : p1221–1227, Jun 1986.
- [111] SUNDE R. A. ; EVENSON J. K. : Serine incorporation into the selenocysteine moiety of glutathione peroxidase. *The Journal of biological chemistry*, 262(2) : p933–937, Jan 1987.
- [112] YUSUPOV M. M. ; YUSUPOVA G. Z. ; BAUCOM A. ; LIEBERMAN K. ; EARNEST T. N. ; CATE J. H. ; NOLLER H. F. : Crystal structure of the ribosome at 5.5 Å resolution. *Science*, 292(5518) : p883–896, May 2001.
- [113] SCHUWIRTH B. S. ; BOROVINSKAYA M. A. ; HAU C. W. ; ZHANG W. ; VILA-SANJURJO A. ; HOLTON J. M. ; CATE J. H. D. : Structures of the bacterial ribosome at 3.5 Å resolution. *Science*, 310(5749) : p827–834, Nov 2005.
- [114] RAMAKRISHNAN V. : Ribosome structure and the mechanism of translation. *Cell*, 108(4) : p557–572, Feb 2002.
- [115] OGLE J. M. ; CARTER A. P. ; RAMAKRISHNAN V. : Insights into the decoding mechanism from recent ribosome structures. *Trends in biochemical sciences*, 28(5) : p259–266, May 2003.
- [116] ALBERTS B. ; JOHNSON A. ; LEWIS J. ; RAFF M. ; ROBERTS K. ; WALTER P. : *Molecular Biology of the Cell, Fourth Edition*. Chap 6, pp. 342–344. Garland, 2002.
- [117] HOAGLAND M. B. ; STEPHENSON M. L. ; SCOTT J. F. ; HECHT L. I. ; ZAMECNIK P. C. : A soluble ribonucleic acid intermediate in protein synthesis. *The Journal of biological chemistry*, 231(1) : p241–257, Mar 1958.
- [118] AMBROGELLY A. ; PALIOURA S. ; SOLL D. : Natural expansion of the genetic code. *Nature chemical biology*, 3(1) : p29–35, Jan 2007.

- [119] BARON C.; WESTHOF E.; BÖCK A. ; GIEGÉ R. : Solution structure of selenocysteine-inserting tRNA(Sec) from *Escherichia coli*. Comparison with canonical tRNA(Ser). *Journal of molecular biology*, 231(2) : p274–292, May 1993.
- [120] BIOU V.; YAREMCHUK A.; TUKALO M. ; CUSACK S. : The 2.9 Å crystal structure of *T. thermophilus* seryl-tRNA synthetase complexed with tRNA(Ser). *Science*, 263(5152) : p1404–1410, Mar 1994.
- [121] CARLSON B. A.; XU X.-M.; KRYUKOV G. V.; RAO M.; BERRY M. J.; GLADYSHEV V. N. ; HATFIELD D. L. : Identification and characterization of phosphoseryl-tRNA[Ser]Sec kinase. *Proceedings of the National Academy of Sciences of the United States of America*, 101(35) : p12848–12853, Aug 2004.
- [122] XU X.-M.; CARLSON B. A.; IRONS R.; MIX H.; ZHONG N.; GLADYSHEV V. N. ; HATFIELD D. L. : Selenophosphate synthetase 2 is essential for selenoprotein biosynthesis. *The Biochemical journal*, 404(1) : p115–120, May 2007.
- [123] RINGQUIST S.; SCHNEIDER D.; GIBSON T.; BARON C.; BÖCK A. ; GOLD L. : Recognition of the mRNA selenocysteine insertion sequence by the specialized translational elongation factor SELB. *Genes & development*, 8(3) : p376–385, Feb 1994.
- [124] ZINONI F.; HEIDER J. ; BÖCK A. : Features of the formate dehydrogenase mRNA necessary for decoding of the UGA codon as selenocysteine. *Proceedings of the National Academy of Sciences of the United States of America*, 87(12) : p4660–4664, Jun 1990.
- [125] HEIDER J.; BARON C. ; BÖCK A. : Coding from a distance : dissection of the mRNA determinants required for the incorporation of selenocysteine into protein. *The EMBO journal*, 11(10) : p3759–3766, Oct 1992.
- [126] CHEN G. F.; FANG L. ; INOUE M. : Effect of the relative position of the UGA codon to the unique secondary structure in the fdhF mRNA on its decoding by selenocysteinyl tRNA in *Escherichia coli*. *The Journal of biological chemistry*, 268(31) : p23128–23131, Nov 1993.
- [127] HÜTTENHOFER A.; HEIDER J. ; BÖCK A. : Interaction of the *Escherichia coli* fdhF mRNA hairpin promoting selenocysteine incorporation with the ribosome. *Nucleic acids research*, 24(20) : p3903–3910, Oct 1996.
- [128] BARON C.; HEIDER J. ; BÖCK A. : Interaction of translation factor SELB with the formate dehydrogenase H selenopolyptide mRNA. *Proceedings of the National Academy of Sciences of the United States of America*, 90(9) : p4181–4185, May 1993.
- [129] ALLMANG C.; KROL A. : Selenoprotein synthesis : UGA does not end the story. *Biochimie*, 88(11) : p1561–1571, Nov 2006.
- [130] ROTHER M.; RESCH A.; WILTING R. ; BÖCK A. : Selenoprotein synthesis in archaea. *BioFactors*, 14(1-4) : p75–83, 2001.
- [131] CABAN K.; KINZY S. A. ; COPELAND P. R. : The L7Ae RNA binding motif is a multifunctional domain required for the ribosome-dependent Sec incorporation activity of Sec insertion sequence binding protein 2. *Molecular and cellular biology*, 27(18) : p6350–6360, Sep 2007.
- [132] CHAVATTE L.; BROWN B. A. ; DRISCOLL D. : Ribosomal protein L30 is a component of the UGA-selenocysteine recoding machinery in eukaryotes. *Nature structural & molecular biology*, 12(5) : p408–416, May 2005.
- [133] KINZY S. A.; CABAN K. ; COPELAND P. R. : Characterization of the SECIS binding protein 2 complex required for the co-translational insertion of selenocysteine in mammals. *Nucleic acids research*, 33(16) : p5172–5180, 2005.
- [134] PAPP L. V.; LU J.; STRIEBEL F.; KENNEDY D.; HOLMGREN A. ; KHANNA K. K. : The redox state of SECIS binding protein 2 controls its localization and selenocysteine incorporation function. *Molecular and cellular biology*, 26(13) : p4895–4910, Jul 2006.
- [135] XU X.-M.; MIX H.; CARLSON B. A.; GRABOWSKI P. J.; GLADYSHEV V. N.; BERRY M. J. ; HATFIELD D. L. : Evidence for direct roles of two additional factors, SECp43 and soluble liver antigen, in the selenoprotein synthesis machinery. *The Journal of biological chemistry*, 280(50) : p41568–41575, Dec 2005.

- 
- [136] CARLSON B. A. ; XU X.-M. ; GLADYSHEV V. N. ; HATFIELD D. L. : Selective rescue of seleno-protein expression in mice lacking a highly specialized methyl group in selenocysteine tRNA. *The Journal of biological chemistry*, 280(7) : p5542–5548, Feb 2005.
- [137] LACOURCIERE G. M. : Selenium is mobilized in vivo from free selenocysteine and is incorporated specifically into formate dehydrogenase H and tRNA nucleosides. *Journal of bacteriology*, 184(7) : p1940–1946, Apr 2002.
- [138] MIHARA H. ; KURIHARA T. ; WATANABE T. ; YOSHIMURA T. ; ESAKI N. : cDNA cloning, purification, and characterization of mouse liver selenocysteine lyase. Candidate for selenium delivery protein in selenoprotein synthesis. *The Journal of biological chemistry*, 275(9) : p6195–6200, Mar 2000.
- [139] MIHARA H. ; MAEDA M. ; FUJII T. ; KURIHARA T. ; HATA Y. ; ESAKI N. : A nifS-like gene, csdB, encodes an Escherichia coli counterpart of mammalian selenocysteine lyase. Gene cloning, purification, characterization and preliminary x-ray crystallographic studies. *The Journal of biological chemistry*, 274(21) : p14768–14772, May 1999.
- [140] LIMA C. D. : Analysis of the E. coli NifS CsdB protein at 2.0 Å reveals the structural basis for perselenide and persulfide intermediate formation. *Journal of molecular biology*, 315(5) : p1199–1208, Feb 2002.
- [141] LACOURCIERE G. M. ; STADTMAN T. C. : Utilization of selenocysteine as a source of selenium for selenophosphate biosynthesis. *BioFactors*, 14(1-4) : p69–74, 2001.
- [142] MIHARA H. ; KATO S.-I. ; LACOURCIERE G. M. ; STADTMAN T. C. ; KENNEDY R. A. J. D. ; KURIHARA T. ; TOKUMOTO U. ; TAKAHASHI Y. ; ESAKI N. : The iscS gene is essential for the biosynthesis of 2-selenouridine in tRNA and the selenocysteine-containing formate dehydrogenase H. *Proceedings of the National Academy of Sciences of the United States of America*, 99(10) : p6679–6683, May 2002.
- [143] PERSSON B. C. ; BÖCK A. ; JÄCKLE H. ; VORBRÜGGEN G. : SelD homolog from Drosophila lacking selenide-dependent monoselenophosphate synthetase activity. *Journal of molecular biology*, 274(2) : p174–180, Nov 1997.
- [144] TAMURA T. ; YAMAMOTO S. ; TAKAHATA M. ; SAKAGUCHI H. ; TANAKA H. ; STADTMAN T. C. ; INAGAKI K. : Selenophosphate synthetase genes from lung adenocarcinoma cells : Sps1 for recycling L-selenocysteine and Sps2 for selenite assimilation. *Proceedings of the National Academy of Sciences of the United States of America*, 101(46) : p16162–16167, Nov 2004.
- [145] FENG L. ; SHEPPARD K. ; NAMGOONG S. ; AMBROGELLY A. ; POLYCARPO C. ; RANDAU L. ; TUMBULA-HANSEN D. ; SÖLL D. : Aminoacyl-tRNA synthesis by pre-translational amino acid modification. *RNA biology*, 1(1) : p16–20, May 2004.
- [146] WILCOX M. ; NIRENBERG M. : Transfer RNA as a cofactor coupling amino acid synthesis with that of protein. *Proceedings of the National Academy of Sciences of the United States of America*, 61(1) : p229–236, Sep 1968.
- [147] CURNOW A. W. ; IBBA M. ; SÖLL D. : tRNA-dependent asparagine formation. *Nature*, 382(6592) : p589–590, Aug 1996.
- [148] BECKER H. D. ; KERN D. : Thermus thermophilus : a link in evolution of the tRNA-dependent amino acid amidation pathways. *Proceedings of the National Academy of Sciences of the United States of America*, 95(22) : p12832–12837, Oct 1998.
- [149] FENG L. ; TUMBULA-HANSEN D. ; TOOGOOD H. ; SÖLL D. : Expanding tRNA recognition of a tRNA synthetase by a single amino acid change. *Proceedings of the National Academy of Sciences of the United States of America*, 100(10) : p5676–5681, May 2003.
- [150] WOESE C. R. ; OLSEN G. J. ; IBBA M. ; SÖLL D. : Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiology and molecular biology reviews : MMBR*, 64(1) : p202–236, Mar 2000.
- [151] HAO B. ; GONG W. ; FERGUSON T. K. ; JAMES C. M. ; KRZYCKI J. A. ; CHAN M. K. : A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science*, 296(5572) : p1462–1466, May 2002.

- [152] GOODCHILD A. ; SAUNDERS N. F. W. ; ERTAN H. ; RAFTERY M. ; GUILHAUS M. ; CURMI P. M. G. ; CAVICCHIOLI R. : A proteomic determination of cold adaptation in the Antarctic archaeon, *Methanococcoides burtonii*. *Molecular microbiology*, 53(1) : p309–321, Jul 2004.
- [153] HERRING S. ; AMBROGELLY A. ; POLYCARPO C. R. ; SÖLL D. : Recognition of pyrrolysine tRNA by the *Desulfitobacterium hafniense* pyrrolysyl-tRNA synthetase. *Nucleic acids research*, 35(4) : p1270–1278, 2007.
- [154] ZHANG Y. ; GLADYSHEV V. N. : High content of proteins containing 21st and 22nd amino acids, selenocysteine and pyrrolysine, in a symbiotic deltaproteobacterium of gutless worm *Olavius algarvensis*. *Nucleic acids research*, 35(15) : p4952–4963, 2007.
- [155] PAUL L. ; FERGUSON J. D. J. ; KRZYCKI J. A. : The trimethylamine methyltransferase gene and multiple dimethylamine methyltransferase genes of *Methanosarcina barkeri* contain in-frame and read-through amber codons. *Journal of bacteriology*, 182(9) : p2520–2529, May 2000.
- [156] JAMES C. M. ; FERGUSON T. K. ; LEYKAM J. F. ; KRZYCKI J. A. : The amber codon in the gene encoding the monomethylamine methyltransferase isolated from *Methanosarcina barkeri* is translated as a sense codon. *The Journal of biological chemistry*, 276(36) : p34252–34258, Sep 2001.
- [157] SRINIVASAN G. ; JAMES C. M. ; KRZYCKI J. A. : Pyrrolysine encoded by UAG in Archaea : charging of a UAG-decoding specialized tRNA. *Science*, 296(5572) : p1459–1462, May 2002.
- [158] SCHIMMEL P. ; BEEBE K. : Molecular biology : genetic code seizes pyrrolysine. *Nature*, 431(7006) : p257–258, Sep 2004.
- [159] KRZYCKI J. A. : The direct genetic encoding of pyrrolysine. *Current opinion in microbiology*, 8(6) : p706–712, Dec 2005.
- [160] THÉOBALD-DIETRICH A. ; GIEGÉ R. ; RUDINGER-THIRION J. : Evidence for the existence in mRNAs of a hairpin element responsible for ribosome dependent pyrrolysine insertion into proteins. *Biochimie*, 87(9-10) : p813–817, Oct 2005.
- [161] POLYCARPO C. R. ; HERRING S. ; BÉRUBÉ A. ; WOOD J. L. ; SÖLL D. ; AMBROGELLY A. : Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase. *FEBS letters*, 580(28-29) : p6695–6700, Dec 2006.
- [162] NAMY O. ; ZHOU Y. ; GUNDLLAPALLI S. ; POLYCARPO C. R. ; DENISE A. ; ROUSSET J.-P. ; SÖLL D. ; AMBROGELLY A. : Adding pyrrolysine to the *Escherichia coli* genetic code. *FEBS letters*, 581(27) : p5282–5288, Nov 2007.
- [163] LONGSTAFF D. G. ; BLIGHT S. K. ; ZHANG L. ; GREEN-CHURCH K. B. ; KRZYCKI J. A. : In vivo contextual requirements for UAG translation as pyrrolysine. *Molecular microbiology*, 63(1) : p229–241, Jan 2007.
- [164] ZHANG Y. ; BARANOV P. V. ; ATKINS J. F. ; GLADYSHEV V. N. : Pyrrolysine and selenocysteine use dissimilar decoding strategies. *The Journal of biological chemistry*, 280(21) : p20740–20751, May 2005.
- [165] ZHANG Y. ; ROMERO H. ; SALINAS G. ; GLADYSHEV V. N. : Dynamic evolution of selenocysteine utilization in bacteria : a balance between selenoprotein loss and evolution of selenocysteine from redox active cysteine residues. *Genome biology*, 7(10) : R94, 2006.
- [166] LONGSTAFF D. G. ; LARUE R. C. ; FAUST J. E. ; MAHAPATRA A. ; ZHANG L. ; GREEN-CHURCH K. B. ; KRZYCKI J. A. : A natural genetic code expansion cassette enables transmissible biosynthesis and genetic encoding of pyrrolysine. *Proceedings of the National Academy of Sciences of the United States of America*, 104(3) : p1021–1026, Jan 2007.
- [167] THÉOBALD-DIETRICH A. ; FRUGIER M. ; GIEGÉ R. ; RUDINGER-THIRION J. : Atypical archaeal tRNA pyrrolysine transcript behaves towards EF-Tu as a typical elongator tRNA. *Nucleic acids research*, 32(3) : p1091–1096, 2004.
- [168] BLIGHT S. ; LARUE R. ; MAHAPATRA A. ; LONGSTAFF D. ; CHANG E. ; ZHAO G. ; KANG P. ; GREEN-CHURCH K. ; CHAN M. ; KRZYCKI J. : Direct charging of tRNA(CUA) with pyrrolysine in vitro and in vivo. *Nature*, 431(7006) : p333–335, Sep 2004.

- 
- [169] AMBROGELLY A.; GUNDLLAPALLI S.; HERRING S.; POLYCARPO C.; FRAUER C. ; SÖLL D. : Pyrrolysine is not hardwired for cotranslational insertion at UAG codons. *Proceedings of the National Academy of Sciences of the United States of America*, 104(9) : p3141–3146, Feb 2007.
- [170] HERRING S.; AMBROGELLY A.; GUNDLLAPALLI S.; O'DONOGHUE P.; POLYCARPO C. R. ; SÖLL D. : The amino-terminal domain of pyrrolysyl-tRNA synthetase is dispensable in vitro but required for in vivo activity. *FEBS letters*, 581(17) : p3197–3203, Jul 2007.
- [171] KAVRAN J. M.; GUNDLLAPALLI S.; O'DONOGHUE P.; ENGLERT M.; SÖLL D. ; STEITZ T. A. : Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proceedings of the National Academy of Sciences of the United States of America*, 104(27) : p11268–11273, Jul 2007.
- [172] HOHN M. J.; PARK H.-S.; O'DONOGHUE P.; SCHNITZBAUER M. ; SOLL D. : Emergence of the universal genetic code imprinted in an RNA record. *Proceedings of the National Academy of Sciences of the United States of America*, 103(48) : p18095–18100, Nov 2006.
- [173] LI T.; GRAHAM D. E.; STATHOPOULOS C.; HANEY P. J.; KIM H. S.; VOTHKNECHT U.; KITABATAKE M.; HONG K. W.; EGGERTSSON G.; CURNOW A. W.; LIN W.; CELIC I.; WHITMAN W. ; SÖLL D. : CysteinyI-tRNA formation : the last puzzle of aminoacyl-tRNA synthesis. *FEBS letters*, 462(3) : p302–306, Dec 1999.
- [174] FUKUNAGA R.; YOKOYAMA S. : Structural insights into the second step of RNA-dependent cysteine biosynthesis in archaea : crystal structure of Sep-tRNA :Cys-tRNA synthase from *Archaeoglobus fulgidus*. *Journal of molecular biology*, 370(1) : p128–141, Jun 2007.
- [175] RUAN B.; NAKANO H.; TANAKA M.; MILLS J. A.; DEVITO J. A.; MIN B.; LOW K. B.; BATTISTA J. R. ; SÖLL D. : CysteinyI-tRNA(Cys) formation in *Methanocaldococcus jannaschii* : the mechanism is still unknown. *Journal of bacteriology*, 186(1) : p8–14, Jan 2004.
- [176] SAUERWALD A.; ZHU W.; MAJOR T. A.; ROY H.; PALIOURA S.; JAHN D.; WHITMAN W. B.; YATES R. J. R.; IBBA M. ; SÖLL D. : RNA-dependent cysteine biosynthesis in archaea. *Science*, 307(5717) : p1969–1972, Mar 2005.
- [177] STATHOPOULOS C.; KIM W.; LI T.; ANDERSON I.; DEUTSCH B.; PALIOURA S.; WHITMAN W. ; SÖLL D. : CysteinyI-tRNA synthetase is not essential for viability of the archaeon *Methanococcus maripaludis*. *Proceedings of the National Academy of Sciences of the United States of America*, 98(25) : p14292–14297, Dec 2001.
- [178] STATHOPOULOS C.; LI T.; LONGMAN R.; VOTHKNECHT U. C.; BECKER H. D.; IBBA M. ; SÖLL D. : One polypeptide with two aminoacyl-tRNA synthetase activities. *Science*, 287(5452) : p479–482, Jan 2000.
- [179] FABREGA C.; FARROW M.; MUKHOPADHYAY B.; DE CRECY-LAGARD V.; ORTIZ A. ; SCHIMMEL P. : An aminoacyl tRNA synthetase whose sequence fits into neither of the two known classes. *Nature*, 411(6833) : p110–114, May 2001.
- [180] AMBROGELLY A.; KAMTEKAR S.; SAUERWALD A.; RUAN B.; TUMBULA-HANSEN D.; KENNEDY D.; AHEL I. ; SÖLL D. : Cys-tRNACys formation and cysteine biosynthesis in methanogenic archaea : two faces of the same problem? *Cellular and molecular life sciences : CMLS*, 61(19-20) : p2437–2445, Oct 2004.
- [181] HELGADÓTTIR S.; ROSAS-SANDOVAL G.; SÖLL D. ; GRAHAM D. E. : Biosynthesis of phosphoserine in the Methanococcales. *Journal of bacteriology*, 189(2) : p575–582, Jan 2007.
- [182] WHITE R. H. : The biosynthesis of cysteine and homocysteine in *Methanococcus jannaschii*. *Biochimica et biophysica acta*, 1624(1-3) : p46–53, Dec 2003.
- [183] KAMTEKAR S.; HOHN M. J.; PARK H.-S.; SCHNITZBAUER M.; SAUERWALD A.; SÖLL D. ; STEITZ T. A. : Toward understanding phosphoseryl-tRNACys formation : the crystal structure of *Methanococcus maripaludis* phosphoseryl-tRNA synthetase. *Proceedings of the National Academy of Sciences of the United States of America*, 104(8) : p2620–2625, Feb 2007.
- [184] FUKUNAGA R.; YOKOYAMA S. : Structural insights into the first step of RNA-dependent cysteine biosynthesis in archaea. *Nature structural & molecular biology*, 14(4) : p272–279, Mar 2007.



- [185] O'DONOGHUE P.; SETHI A.; WOESE C. R. ; LUTHEY-SCHULTEN Z. A. : The evolutionary history of Cys-tRNACys formation. *Proceedings of the National Academy of Sciences of the United States of America*, 102(52) : p19003–19008, Dec 2005.
- [186] NAUSER T.; DOCKHEER S.; KISSNER R. ; KOPPENOL W. H. : Catalysis of electron transfer by selenocysteine. *Biochemistry*, 45(19) : p6038–6043, May 2006.
- [187] JACOB C.; GILES G. I.; GILES N. M. ; SIES H. : Sulfur and selenium : the role of oxidation state in protein structure and function. *Angewandte Chemie (International ed. in English)*, 42(39) : p4742–4758, Oct 2003.
- [188] WESSJOHANN L. A.; SCHNEIDER A.; ABBAS M. ; BRANDT W. : Selenium in chemistry and biochemistry in comparison to sulfur. *Biological chemistry*, 388(10) : p997–1006, Oct 2007.
- [189] CARDEY B.; ENESCU M. : Selenocysteine versus cysteine reactivity : a theoretical study of their oxidation by hydrogen peroxide. *The journal of physical chemistry. A*, 111(4) : p673–678, Feb 2007.
- [190] BOYINGTON J. C.; GLADYSHEV V. N.; KHANGULOV S. V.; STADTMAN T. C. ; SUN P. D. : Crystal structure of formate dehydrogenase H : catalysis involving Mo, molybdopterin, selenocysteine, and an Fe<sub>4</sub>S<sub>4</sub> cluster. *Science*, 275(5304) : p1305–1308, Feb 1997.
- [191] AXLEY M. J.; BÖCK A. ; STADTMAN T. C. : Catalytic properties of an Escherichia coli formate dehydrogenase mutant in which sulfur replaces selenium. *Proceedings of the National Academy of Sciences of the United States of America*, 88(19) : p8450–8454, Oct 1991.
- [192] KHANGULOV S. V.; GLADYSHEV V. N.; DISMUKES G. C. ; STADTMAN T. C. : Selenium-containing formate dehydrogenase H from Escherichia coli : a molybdopterin enzyme that catalyzes formate oxidation without oxygen transfer. *Biochemistry*, 37(10) : p3518–3528, Mar 1998.
- [193] KIM H.-Y.; GLADYSHEV V. N. : Methionine sulfoxide reductases : selenoprotein forms and roles in antioxidant protein repair in mammals. *The Biochemical journal*, 407(3) : p321–329, Nov 2007.
- [194] KIM H.-Y.; FOMENKO D. E.; YOON Y.-E. ; GLADYSHEV V. N. : Catalytic advantages provided by selenocysteine in methionine-S-sulfoxide reductases. *Biochemistry*, 45(46) : p13697–13704, Nov 2006.
- [195] STADTMAN T. C. : Selenoproteins—tracing the role of a trace element in protein function. *PLoS biology*, 3(12) : e421, Dec 2005.
- [196] KIM H.-Y.; GLADYSHEV V. N. : Methionine sulfoxide reduction in mammals : characterization of methionine-R-sulfoxide reductases. *Molecular biology of the cell*, 15(3) : p1055–1064, Mar 2004.
- [197] KIM H.-Y.; GLADYSHEV V. N. : Different catalytic mechanisms in mammalian selenocysteine- and cysteine-containing methionine-R-sulfoxide reductases. *PLoS biology*, 3(12) : e375, Dec 2005.
- [198] LEE S. R.; BAR-NOY S.; KWON J.; LEVINE R. L.; STADTMAN T. C. ; RHEE S. G. : Mammalian thioredoxin reductase : oxidation of the C-terminal cysteine/selenocysteine active site forms a thioselenide, and replacement of selenium with sulfur markedly reduces catalytic activity. *Proceedings of the National Academy of Sciences of the United States of America*, 97(6) : p2521–2526, Mar 2000.
- [199] ZHONG L.; HOLMGREN A. : Essential role of selenium in the catalytic activities of mammalian thioredoxin reductase revealed by characterization of recombinant enzymes with selenocysteine mutations. *The Journal of biological chemistry*, 275(24) : p18121–18128, Jun 2000.
- [200] SERVE O.; DESMOLAIZE B.; TERCINIER P. ; GAURON G. : Groquick's death, or When globalisation means culture uniformity. *Le journal de Mickey*, 666(3) : p51–69, 1999.
- [201] BAR-NOY S.; GORLATOV S. N. ; STADTMAN T. C. : Overexpression of wild type and SeCys/Cys mutant of human thioredoxin reductase in E. coli : the role of selenocysteine in the catalytic activity. *Free radical biology & medicine*, 30(1) : p51–61, Jan 2001.
- [202] FUJIWARA N.; FUJII T.; FUJII J. ; TANIGUCHI N. : Roles of N-terminal active cysteines and C-terminal cysteine-selenocysteine in the catalytic mechanism of mammalian thioredoxin reductase. *Journal of biochemistry*, 129(5) : p803–812, May 2001.

- 
- [203] BRANDT W.; WESSJOHANN L. A. : The functional role of selenocysteine (Sec) in the catalysis mechanism of large thioredoxin reductases : proposition of a swapping catalytic triad including a Sec-His-Glu state. *Chembiochem*, 6(2) : p386–394, Feb 2005.
- [204] KUIPER G. G. J. M.; KLOOTWIJK W.; VISSER T. J. : Substitution of cysteine for selenocysteine in the catalytic center of type III iodothyronine deiodinase reduces catalytic efficiency and alters substrate preference. *Endocrinology*, 144(6) : p2505–2513, Jun 2003.
- [205] BERRY M. J.; MAIA A. L.; KIEFFER J. D.; HARNEY J. W.; LARSEN P. R. : Substitution of cysteine for selenocysteine in type I iodothyronine deiodinase reduces the catalytic efficiency of the protein but enhances its translation. *Endocrinology*, 131(4) : p1848–1852, Oct 1992.
- [206] BUETTNER C.; HARNEY J. W.; LARSEN P. R. : The role of selenocysteine 133 in catalysis by the human type 2 iodothyronine deiodinase. *Endocrinology*, 141(12) : p4606–4612, Dec 2000.
- [207] KIM I. Y.; GUIMARÃES M. J.; ZLOTNIK A.; BAZAN J. F.; STADTMAN T. C. : Fetal mouse selenophosphate synthetase 2 (SPS2) : characterization of the cysteine mutant form overproduced in a baculovirus-insect cell system. *Proceedings of the National Academy of Sciences of the United States of America*, 94(2) : p418–421, Jan 1997.
- [208] LACOURCIERE G. M.; STADTMAN T. C. : Catalytic properties of selenophosphate synthetases : comparison of the selenocysteine-containing enzyme from *Haemophilus influenzae* with the corresponding cysteine-containing enzyme from *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, 96(1) : p44–48, Jan 1999.
- [209] VERES Z.; KIM I. Y.; SCHOLZ T. D.; STADTMAN T. C. : Selenophosphate synthetase. Enzyme properties and catalytic reaction. *The Journal of biological chemistry*, 269(14) : p10597–10603, Apr 1994.
- [210] GARCIN E.; VERNEDE X.; HATCHIKIAN E. C.; VOLBEDA A.; FREY M.; FONTECILLA-CAMPS J. C. : The crystal structure of a reduced [NiFeSe] hydrogenase provides an image of the activated catalytic center. *Structure*, 7(5) : p557–566, May 1999.
- [211] FAUQUE G.; PECK J. H. D.; MOURA J. J.; HUYNH B. H.; BERLIER Y.; DERVARTANIAN D. V.; TEIXEIRA M.; PRZYBYLA A. E.; LESPINAT P. A.; MOURA I. : The three classes of hydrogenases from sulfate-reducing bacteria of the genus *Desulfovibrio*. *FEMS microbiology reviews*, 4(4) : p299–344, Dec 1988.
- [212] ARKOWITZ R. A.; ABELES R. H. : Mechanism of action of clostridial glycine reductase : isolation and characterization of a covalent acetyl enzyme intermediate. *Biochemistry*, 30(16) : p4090–4097, Apr 1991.
- [213] TASKOV K.; CHAPPLE C.; KRYUKOV G. V.; CASTELLANO S.; LOBANOV A. V.; KOROTKOV K. V.; GUIGÓ R.; GLADYSHEV V. N. : Nematode selenoproteome : the use of the selenocysteine insertion system to decode one codon in an animal genome? *Nucleic acids research*, 33(7) : p2227–2238, 2005.
- [214] LOBANOV A.; HATFIELD D.; GLADYSHEV V. : Reduced reliance on the trace element selenium during evolution of mammals. *Genome biology*, 9(3) : R62, Mar 2008.
- [215] HATFIELD D. L.; CARLSON B. A.; XU X.-M.; MIX H.; GLADYSHEV V. N. : Selenocysteine incorporation machinery and the role of selenoproteins in development and health. *Progress in nucleic acid research and molecular biology*, 81 : p97–142, 2006.
- [216] BOCK A. : *Selenium metabolism in bacteria*. In “*Selenium : Its Molecular Biology and Role in Human Health*”. Chap 2, pp. 7–22. D. L. Hatfield Ed., Kluwer Academic Publishers, Norwell MA, USA, 2001.
- [217] GLADYSHEV V. N.; KRYUKOV G. V. : Evolution of selenocysteine-containing proteins : significance of identification and functional characterization of selenoproteins. *BioFactors*, 14(1-4) : p87–92, 2001.
- [218] ARNÉR E. S.; SARIOGLU H.; LOTTSPREICH F.; HOLMGREN A.; BÖCK A. : High-level expression in *Escherichia coli* of selenocysteine-containing rat thioredoxin reductase utilizing gene fusions with engineered bacterial-type SECIS elements and co-expression with the selA, selB and selC genes. *Journal of molecular biology*, 292(5) : p1003–1016, Oct 1999.

- [219] SU D.; LI Y. ; GLADYSHEV V. N. : Selenocysteine insertion directed by the 3'-UTR SECIS element in Escherichia coli. *Nucleic acids research*, 33(8) : p2486–2492, 2005.
- [220] MIX H.; LOBANOV A. V. ; GLADYSHEV V. N. : SECIS elements in the coding regions of selenoprotein transcripts are functional in higher eukaryotes. *Nucleic acids research*, 35(2) : p414–423, 2007.
- [221] SOLER N. : *Etude structurale de la protéine SelB, impliquée dans l'incorporation des sélénocystéines lors de la synthèse protéique bactérienne*. 216 pages. Thèse de Doctorat, Biochimie, Biologie structurale, Université Paris 11, Jul 2007.
- [222] SCHÖN A.; BÖCK A.; OTT G.; SPRINZL M. ; SÖLL D. : The selenocysteine-inserting opal suppressor serine tRNA from E. coli is highly unusual in structure and modification. *Nucleic acids research*, 17(18) : p7159–7165, Sep 1989.
- [223] BARON C.; HEIDER J. ; BÖCK A. : Mutagenesis of selC, the gene for the selenocysteine-inserting tRNA-species in E. coli : effects on in vivo function. *Nucleic acids research*, 18(23) : p6761–6766, Dec 1990.
- [224] HUBERT N.; STURCHLER C.; WESTHOF E.; CARBON P. ; KROL A. : The 9/4 secondary structure of eukaryotic selenocysteine tRNA : more pieces of evidence. *RNA*, 4(9) : p1029–1033, Sep 1998.
- [225] MIZUTANI T.; GOTO C. : Eukaryotic selenocysteine tRNA has the 9/4 secondary structure. *FEBS letters*, 466(2-3) : p359–362, Jan 2000.
- [226] STURCHLER C.; WESTHOF E.; CARBON P. ; KROL A. : Unique secondary and tertiary structural features of the eucaryotic selenocysteine tRNA(Sec). *Nucleic acids research*, 21(5) : p1073–1079, Mar 1993.
- [227] TORMAY P.; WILTING R.; HEIDER J. ; BÖCK A. : Genes coding for the selenocysteine-inserting tRNA species from Desulfomicrobium baculatum and Clostridium thermoaceticum : structural and evolutionary implications. *Journal of bacteriology*, 176(5) : p1268–1274, Mar 1994.
- [228] BARON C.; STURCHLER C.; WU X. Q.; GROSS H. J.; KROL A. ; BÖCK A. : Eukaryotic selenocysteine inserting tRNA species support selenoprotein synthesis in Escherichia coli. *Nucleic acids research*, 22(12) : p2228–2233, Jun 1994.
- [229] COMMANS S.; BÖCK A. : Selenocysteine inserting tRNAs : an overview. *FEMS microbiology reviews*, 23(3) : p335–351, Jun 1999.
- [230] BARON C.; BÖCK A. : The length of the aminoacyl-acceptor stem of the selenocysteine-specific tRNA(Sec) of Escherichia coli is the determinant for binding to elongation factors SELB or Tu. *The Journal of biological chemistry*, 266(30) : p20375–20379, Oct 1991.
- [231] FÖRSTER C.; OTT G.; FÖRCHHAMMER K. ; SPRINZL M. : Interaction of a selenocysteine-incorporating tRNA with elongation factor Tu from E.coli. *Nucleic acids research*, 18(3) : p487–491, Feb 1990.
- [232] RUDINGER J.; HILLENBRANDT R.; SPRINZL M. ; GIEGÉ R. : Antideterminants present in minihelix(Sec) hinder its recognition by prokaryotic elongation factor Tu. *The EMBO journal*, 15(3) : p650–657, Feb 1996.
- [233] SHERRER R. L.; HO J. M. L. ; SÖLL D. : Divergence of selenocysteine tRNA recognition by archaeal and eukaryotic O-phosphoseryl-tRNA<sub>Sec</sub> kinase. *Nucleic acids research*, 36(6) : p1871–1880, Apr 2008.
- [234] MIZUTANI T.; KANAYA K.; IKEDA S.; FUJIWARA T.; YAMADA K. ; TOTSUKA T. : The dual identities of mammalian tRNA(Sec) for SerRS and selenocysteine synthase. *Molecular biology reports*, 25(4) : p211–216, Nov 1998.
- [235] ZHOU X.; PARK S. I.; MOUSTAFA M. E.; CARLSON B. A.; CRAIN P. F.; DIAMOND A. M.; HATFIELD D. L. ; LEE B. J. : Selenium metabolism in Drosophila. Characterization of the selenocysteine tRNA population. *The Journal of biological chemistry*, 274(26) : p18729–18734, Jun 1999.

- 
- [236] WARNER G. J.; BERRY M. J.; MOUSTAFA M. E.; CARLSON B. A.; HATFIELD D. L. ; FAUST J. R. : Inhibition of selenoprotein synthesis by selenocysteine tRNA[Ser]Sec lacking isopentenyladenosine. *The Journal of biological chemistry*, 275(36) : p28110–28119, Sep 2000.
- [237] MOUSTAFA M. E.; CARLSON B. A.; EL-SAADANI M. A.; KRYUKOV G. V.; SUN Q. A.; HARNEY J. W.; HILL K. E.; COMBS G. F.; FEIGENBAUM L.; MANSUR D. B.; BURK R. F.; BERRY M. J.; DIAMOND A. M.; LEE B. J.; GLADYSHEV V. N. ; HATFIELD D. L. : Selective inhibition of selenocysteine tRNA maturation and selenoprotein synthesis in transgenic mice expressing isopentenyladenosine-deficient selenocysteine tRNA. *Molecular and cellular biology*, 21(11) : p3840–3852, Jun 2001.
- [238] STURCHLER C.; LESCURE A.; KEITH G.; CARBON P. ; KROL A. : Base modification pattern at the wobble position of *Xenopus* selenocysteine tRNA(Sec). *Nucleic acids research*, 22(8) : p1354–1358, Apr 1994.
- [239] HATFIELD D.; LEE B. J.; HAMPTON L. ; DIAMOND A. M. : Selenium induces changes in the selenocysteine tRNA[Ser]Sec population in mammalian cells. *Nucleic acids research*, 19(4) : p939–943, Feb 1991.
- [240] JAMESON R. R.; DIAMOND A. M. : A regulatory role for Sec tRNA[Ser]Sec in selenoprotein synthesis. *RNA*, 10(7) : p1142–1152, Jul 2004.
- [241] KIM L. K.; MATSUFUJI T.; MATSUFUJI S.; CARLSON B. A.; KIM S. S.; HATFIELD D. L. ; LEE B. J. : Methylation of the ribosyl moiety at position 34 of selenocysteine tRNA[Ser]Sec is governed by both primary and tertiary structure. *RNA*, 6(9) : p1306–1315, Sep 2000.
- [242] YUAN J.; PALIOURA S.; SALAZAR J. C.; SU D.; O'DONOGHUE P.; HOHN M. J.; CARDOSO A. M.; WHITMAN W. B. ; SÖLL D. : RNA-dependent conversion of phosphoserine forms selenocysteine in eukaryotes and archaea. *Proceedings of the National Academy of Sciences of the United States of America*, 103(50) : p18923–18927, Dec 2006.
- [243] RAO M.; CARLSON B. A.; NOVOSELOV S. V.; WEEKS D. P.; GLADYSHEV V. N. ; HATFIELD D. L. : *Chlamydomonas reinhardtii* selenocysteine tRNA[Ser]Sec. *RNA*, 9(8) : p923–930, Aug 2003.
- [244] LEINFELDER W.; FORCHHAMMER K.; VEPREK B.; ZEHELEIN E. ; BÖCK A. : In vitro synthesis of selenocysteinyl-tRNA(UCA) from seryl-tRNA(UCA) : involvement and characterization of the selD gene product. *Proceedings of the National Academy of Sciences of the United States of America*, 87(2) : p543–547, Jan 1990.
- [245] GLASS R. S.; SINGH W. P.; JUNG W.; VERES Z.; SCHOLZ T. D. ; STADTMAN T. C. : Mono-selenophosphate : synthesis, characterization, and identity with the prokaryotic biological selenium donor, compound SePX. *Biochemistry*, 32(47) : p12555–12559, Nov 1993.
- [246] GLASS R. S.; STADTMAN T. C. : Selenophosphate. *Methods in enzymology*, 252 : p309–315, 1995.
- [247] MIZUTANI T.; KANAYA K. ; TANABE K. : Selenophosphate as a substrate for mammalian selenocysteine synthase, its stability and toxicity. *BioFactors*, 9(1) : p27–36, 1999.
- [248] KIM I. Y.; VERES Z. ; STADTMAN T. C. : *Escherichia coli* mutant SELD enzymes. The cysteine 17 residue is essential for selenophosphate formation from ATP and selenide. *The Journal of biological chemistry*, 267(27) : p19650–19654, Sep 1992.
- [249] KIM I. Y.; VERES Z. ; STADTMAN T. C. : Biochemical analysis of *Escherichia coli* selenophosphate synthetase mutants. Lysine 20 is essential for catalytic activity and cysteine 17/19 for 8-azido-ATP derivatization. *The Journal of biological chemistry*, 268(36) : p27020–27025, Dec 1993.
- [250] FORCHHAMMER K.; BÖCK A. : Selenocysteine synthase from *Escherichia coli*. Analysis of the reaction sequence. *The Journal of biological chemistry*, 266(10) : p6324–6328, Apr 1991.
- [251] FORCHHAMMER K.; LEINFELDER W.; BOESMILLER K.; VEPREK B. ; BÖCK A. : Selenocysteine synthase from *Escherichia coli*. Nucleotide sequence of the gene (selA) and purification of the protein. *The Journal of biological chemistry*, 266(10) : p6318–6323, Apr 1991.
- [252] SAWERS G.; HEIDER J.; ZEHELEIN E. ; BÖCK A. : Expression and operon structure of the sel genes of *Escherichia coli* and identification of a third selenium-containing formate dehydrogenase isoenzyme. *Journal of bacteriology*, 173(16) : p4983–4993, Aug 1991.

- [253] FORCHHAMMER K. ; BOESMILLER K. ; BÖCK A. : The function of selenocysteine synthase and SELB in the synthesis and incorporation of selenocysteine. *Biochimie*, 73(12) : p1481–1486, Dec 1991.
- [254] ENGELHARDT H. ; FORCHHAMMER K. ; MÜLLER S. ; GOLDIE K. N. ; BÖCK A. : Structure of selenocysteine synthase from *Escherichia coli* and location of tRNA in the seryl-tRNA(sec)-enzyme complex. *Molecular microbiology*, 6(23) : p3461–3467, Dec 1992.
- [255] TORMAY P. ; WILTING R. ; LOTTSPREICH F. ; MEHTA P. K. ; CHRISTEN P. ; BÖCK A. : Bacterial selenocysteine synthase—structural and functional properties. *European journal of biochemistry / FEBS*, 254(3) : p655–661, Jun 1998.
- [256] FISCHER N. ; PALESKAVA A. ; GROMADSKI K. B. ; KONEVEGA A. L. ; WAHL M. C. ; STARK H. ; RODNINA M. V. : Towards understanding selenocysteine incorporation into bacterial proteins. *Biological chemistry*, 388(10) : p1061–1067, Oct 2007.
- [257] PERCUDANI R. ; PERACCHI A. : A genomic overview of pyridoxal-phosphate-dependent enzymes. *EMBO reports*, 4(9) : p850–854, Sep 2003.
- [258] MEHTA P. K. ; CHRISTEN P. : The molecular evolution of pyridoxal-5'-phosphate-dependent enzymes. *Advances in enzymology and related areas of molecular biology*, 74 : p129–184, 2000.
- [259] XU X.-M. ; CARLSON B. A. ; MIX H. ; ZHANG Y. ; SAIRA K. ; GLASS R. S. ; BERRY M. J. ; GLADYSHEV V. N. ; HATFIELD D. L. : Biosynthesis of selenocysteine on its tRNA in eukaryotes. *PLoS biology*, 5(1) : e4, Jan 2007.
- [260] ARAISO Y. ; PALIOURA S. ; ISHITANI R. ; SHERRER R. L. ; O'DONOGHUE P. ; YUAN J. ; OSHIKANE H. ; DOMAE N. ; DEFRANCO J. ; SÖLL D. ; NUREKI O. : Structural insights into RNA-dependent eukaryal and archaeal selenocysteine formation. *Nucleic acids research*, 36(4) : p1187–1199, Mar 2008.
- [261] PAIARDINI A. ; BOSSA F. ; PASCARELLA S. : Evolutionarily conserved regions and hydrophobic contacts at the superfamily level : The case of the fold-type I, pyridoxal-5'-phosphate-dependent enzymes. *Protein science*, 13(11) : p2992–3005, Nov 2004.
- [262] HAYASHI H. : Pyridoxal enzymes : mechanistic diversity and uniformity. *Journal of biochemistry*, 118(3) : p463–473, Sep 1995.
- [263] VEDERAS J. ; FLOSS H. : Stereochemistry of Pyridoxal Phosphate Catalyzed Enzyme. *Accounts of Chemical Research*, 13(12) : p455–463, 1980.
- [264] ALEXANDER F. W. ; SANDMEIER E. ; MEHTA P. K. ; CHRISTEN P. : Evolutionary relationships among pyridoxal-5'-phosphate-dependent enzymes. Regio-specific alpha, beta and gamma families. *European journal of biochemistry / FEBS*, 219(3) : p953–960, Feb 1994.
- [265] KAISER J. T. ; GROMADSKI K. ; ROTHER M. ; ENGELHARDT H. ; RODNINA M. V. ; WAHL M. C. : Structural and functional investigation of a putative archaeal selenocysteine synthase. *Biochemistry*, 44(40) : p13315–13327, Oct 2005.
- [266] GANICHKIN O. M. ; XU X.-M. ; CARLSON B. A. ; MIX H. ; HATFIELD D. L. ; GLADYSHEV V. N. ; WAHL M. C. : Structure and catalytic mechanism of eukaryotic selenocysteine synthase. *The Journal of biological chemistry*, 283(9) : p5849–5865, Feb 2008.
- [267] KATTI S. K. ; KATZ B. A. ; WYCKOFF H. W. : Crystal structure of muconolactone isomerase at 3.3 Å resolution. *Journal of molecular biology*, 205(3) : p557–571, Feb 1989.
- [268] VONCK J. ; ARFMAN N. ; DE VRIES G. E. ; VAN BEEUMEN J. ; VAN BRUGGEN E. F. ; DIJKHUIZEN L. : Electron microscopic analysis and biochemical characterization of a novel methanol dehydrogenase from the thermotolerant *Bacillus* sp. C1. *The Journal of biological chemistry*, 266(6) : p3949–3954, Feb 1991.
- [269] HAMIAUX C. ; PRANGÉ T. ; RIÈS-KAUTT M. ; DUCRUX A. ; LAFONT S. ; ASTIER J. P. ; VEESLER S. : The decameric structure of bovine pancreatic trypsin inhibitor (BPTI) crystallized from thiocyanate at 2.7 Å resolution. *Acta crystallographica. Section D, Biological crystallography*, 55(Pt 1) : p103–113, Jan 1999.

- 
- [270] WALSH M. A. ; OTWINOWSKI Z. ; PERRAKIS A. ; ANDERSON P. M. ; JOACHIMIAK A. : Structure of cyanase reveals that a novel dimeric and decameric arrangement of subunits is required for formation of the enzyme active site. *Structure*, 8(5) : p505–514, May 2000.
- [271] SCHRÖDER E. ; LITTLECHILD J. A. ; LEBEDEV A. A. ; ERRINGTON N. ; VAGIN A. A. ; ISUPOV M. N. : Crystal structure of decameric 2-Cys peroxiredoxin from human erythrocytes at 1.7 Å resolution. *Structure*, 8(6) : p605–615, Jun 2000.
- [272] THORELL S. ; SCHÜRMAN M. ; SPRENGER G. A. ; SCHNEIDER G. : Crystal structure of decameric fructose-6-phosphate aldolase from *Escherichia coli* reveals inter-subunit helix swapping as a structural basis for assembly differences in the transaldolase family. *Journal of molecular biology*, 319(1) : p161–171, May 2002.
- [273] MAITA N. ; OKADA K. ; HATAKEYAMA K. ; HAKOSHIMA T. : Crystal structure of the stimulatory complex of GTP cyclohydrolase I and its feedback regulatory protein GFRP. *Proceedings of the National Academy of Sciences of the United States of America*, 99(3) : p1212–1217, Feb 2002.
- [274] CHAUDHURI B. N. ; SAWAYA M. R. ; KIM C. Y. ; WALDO G. S. ; PARK M. S. ; TERWILLIGER T. C. ; YEATES T. O. : The crystal structure of the first enzyme in the pantothenate biosynthetic pathway, ketopantoate hydroxymethyltransferase, from *M. tuberculosis*. *Structure*, 11(7) : p753–764, Jul 2003.
- [275] WALKER J. R. ; NAGAR B. ; YOUNG N. M. ; HIRAMA T. ; RINI J. M. : X-ray crystal structure of a galactose-specific C-type lectin possessing a novel decameric quaternary structure. *Biochemistry*, 43(13) : p3783–3792, Apr 2004.
- [276] KLINKE S. ; ZYLBERMAN V. ; VEGA D. R. ; GUIMARÃES B. G. ; BRADEN B. C. ; GOLDBAUM F. A. : Crystallographic studies on decameric *Brucella* spp. Lumazine synthase : a novel quaternary arrangement evolved for a new function ? *Journal of molecular biology*, 353(1) : p124–137, Oct 2005.
- [277] WOOD Z. A. ; SCHRÖDER E. ; ROBIN HARRIS J. ; POOLE L. B. : Structure, mechanism and regulation of peroxiredoxins. *Trends in biochemical sciences*, 28(1) : p32–40, Jan 2003.



## Deuxième partie





La partie bibliographique précédente, recueil de connaissances actuelles sur la sélénocystéine, sur les mécanismes de sa biosynthèse et de son incorporation dans les sélénoprotéines, et sur des domaines d'intérêt attenants à son étude, n'a pu être construite que par l'accumulation, pendant presque trois ans, de données et de réflexions sur le sujet.

La situation était bien différente en septembre 2005, lorsque le Docteur Dominique Fourmy accepta de proposer ce sujet de biologie moléculaire et structurale au chimiste cherchant à changer de thématique que j'étais. Le cheminement entre hier et aujourd'hui est pavé d'une double évolution : celle des connaissances dans ce domaine de recherche mouvant, et mes propres efforts d'adaptation au monde de la biologie, à ses techniques et à ses problématiques.

Ainsi depuis mes débuts dans la thématique, les connaissances disponibles autour de la biosynthèse du 21<sup>ième</sup> acide aminé se sont considérablement étendues : découverte des protéines "SelA-like" archaïques et de la résolution de la structure cristallographique de l'une d'entre elle<sup>[II.1]</sup> en octobre 2005 ; découverte de l'identité des SepSecS des eucaryotes et des archaïques<sup>[II.2]</sup> en décembre 2006 ; et résolution de leurs structures<sup>[II.3,II.4]</sup> en décembre 2007.

Ces avancées importantes ne sont que des exemples dans l'explosion des connaissances que vit ce domaine de recherche. L'intérêt du travail de recherche de notre laboratoire n'en est que plus grand.

L'objectif structural de mon travail de thèse reste encore aujourd'hui la pièce manquante du puzzle qui permettrait d'avoir une vision globale et peut-être une piste vers la compréhension de l'émergence des systèmes de biosynthèse de la sélénocystéine.

L'intérêt de connaître la structure de la sélénocystéine synthase bactérienne n'est pas nouveau, et des tentatives ont été ou sont en cours à travers le monde pour la résoudre. Bien que ce ne soit plus le cas aujourd'hui, cette protéine faisait partie des candidats aux projets de génomique structurale à grand échelle et à haut débit outre-Atlantique. Un intéressant projet étranger actuel est basé sur la collaboration entre l'équipe de Marina V. Rodnina (Institute of Physical Biochemistry, Witten/Herdecke Universität) et l'équipe de cryomicroscopie électronique tridimensionnelle de Holger Stark (Max Planck Institute for Biophysical Chemistry, Göttingen), pour obtenir une structure de SelA à moyenne résolution, grâce aux avancées de la cryomicroscopie électronique.<sup>[II.5]</sup>

Pour notre étude, nous nous sommes intéressés aux couples SelA / ARN<sub>t</sub><sup>Sec</sup> des organismes éloignés *Escherichia coli* et *Moorella thermoacetica*, pour en comparer le fonctionnement et accéder aux données structurales de l'un d'entre eux.

Le premier chapitre de cette 2<sup>ème</sup> partie (chapitre 3) retranscrit, dans un ordre thématique, le cheminement expérimental effectué dans le but premier d'obtenir la structure cristallographique d'une sélénocystéine synthase bactérienne, en insistant sur les améliorations successives apportées à la méthodologie.

Le chapitre suivant (chapitre 4) présente les différentes caractérisations biochimiques du fonctionnement de l'enzyme, avec son partenaire  $\text{ARN}_t^{\text{Sec}}$  ou  $\text{Ser-ARN}_t^{\text{Sec}}$ .

Enfin, le dernier chapitre (chapitre 5) sera une discussion sur la portée des nouveaux éléments produits par les présentes études expérimentales, l'apport de données bio-informatiques et une réflexion sur le fonctionnement et la phylogénie du couple protéo-nucléique sujet de notre étude.

# Chapitre 3

## Vers l'obtention de cristaux du complexe SelA / ARN<sub>t</sub><sup>Sec</sup>

### Sommaire

---

<b>3.1</b>	<b>Choix des partenaires . . . . .</b>	<b>132</b>
3.1.1	Les ARN <sub>t</sub> <sup>Sec</sup> de <i>E. coli</i> et <i>M. thermoacetica</i> . . . . .	133
3.1.2	Caractéristiques des sélénocystéine synthases SelA . . . . .	134
<b>3.2</b>	<b>Obtention des ARN<sub>t</sub><sup>Sec</sup> . . . . .</b>	<b>135</b>
3.2.1	Au départ, <i>selC</i> ... . . . .	135
3.2.2	La transcription . . . . .	139
3.2.3	La charge en sérine . . . . .	140
3.2.4	... au final, le séryl-ARN <sub>t</sub> <sup>Sec</sup> . . . . .	142
<b>3.3</b>	<b>Obtention des sélénocystéine synthases . . . . .</b>	<b>144</b>
3.3.1	Clonage et vecteur d'expression . . . . .	144
3.3.2	Production et récolte . . . . .	144
3.3.3	Purification des constructions protéiques . . . . .	145
3.3.3.1	Méthode initiale . . . . .	145
3.3.3.2	Activité RNase des échantillons . . . . .	146
3.3.3.3	Stabilité, propriétés des protéines . . . . .	147
3.3.3.4	Evolution du protocole de purification . . . . .	148
3.3.3.5	L'ajout de cofacteur . . . . .	151
3.3.3.6	Conclusions sur la purification de SelA . . . . .	151
<b>3.4</b>	<b>Formation du complexe . . . . .</b>	<b>153</b>
3.4.1	Mise en présence des partenaires . . . . .	153
3.4.2	Purification du complexe . . . . .	154
3.4.3	Essais de production directe du complexe . . . . .	155
<b>3.5</b>	<b>Cristallogénèse . . . . .</b>	<b>156</b>
3.5.1	Techniques employées . . . . .	157
3.5.2	Carnet de bord de la cristallogénèse . . . . .	160
3.5.3	Cristaux et reproductibilité . . . . .	169

---

D'un point de vue stratégique, notre démarche est simple : le but premier de ce travail est de réussir à produire l'intermédiaire stable (2) du schéma réactionnel de la page 95 : l'aminocrylyl-ARN<sub>t</sub><sup>Sec</sup> fixé au cofacteur PLP et complexé à SelA.

Cette espèce est la plus intéressante structuralement, car elle permet d'obtenir des informations mécanistiques sur la réaction entre le séryl-ARN<sub>t</sub><sup>Sec</sup> et SelA en plus des informations structurales sur les deux partenaires.

Comme les enzymes SelA d'*Escherichia coli* et de *Moorella thermoacetica* sont compatibles avec les ARN<sub>t</sub><sup>Sec</sup> des deux espèces, quatre types de complexes sont possibles.

Ce chapitre présente comment la production des partenaires, leur étude et la formation de certains de ces complexes, m'a permis d'aboutir à la cristallisation de l'un d'entre eux.

### 3.1 Choix des partenaires

Que ce soit dans un but structural ou d'étude biochimique, le choix des éléments d'étude et la connaissance de leurs caractéristiques sont primordiaux.

Les protéines et ARN employés dans notre travail sont issus des génomes des organismes *Escherichia coli* et *Moorella thermoacetica*.

Ces deux organismes sont des représentants des eubactéries, très éloignés phylogéniquement, *Moorella thermoacetica* appartenant à la famille des Thermoanaerobacteriaceae dans le phyla des Firmicutes/Clostridia, et *Escherichia coli* étant une Enterobacteriaceae du groupe des  $\gamma$ -protéobactéries. L'arbre phylogénique de la figure 1.35, page 75, peut permettre de replacer ces organismes l'un par rapport à l'autre.

En raison de la distance entre ces organismes, — et entre leurs sélénocystéine synthases SelA (voir l'annexe A.3, page 232) — les phénomènes et mécanismes observés dans le fonctionnement de leurs systèmes de biosynthèse de la sélénocystéine peuvent revêtir une portée générale, car ils ne sont pas la spécificité d'un phyla ou d'une famille d'eubactéries.

Ces deux organismes possèdent la capacité de synthétiser des sélénoprotéines, propriété essentielle à notre étude, et leurs génomes ont été totalement séquencés.

Les sélénoprotéines présentes dans le génome de *Moorella thermoacetica* sont au nombre de 6 : il s'agit d'une sélénophosphate synthétase (SelD), deux exemplaires de la formate déshydrogénase, sous-unité A (FdhA), une hydrogénase F420-réductrice, sous-unité  $\delta$  (FrhD), une hétérodisulfure réductase, sous-unité A (HdrA) et une glutarédoxine.

Seulement 3 sélénoprotéines ont été trouvées chez *Escherichia coli*. Il s'agit des isoenzymes suivantes : la formate déshydrogénase H (FdhH), une formate déshydrogénase nitrate-inductible (FdhN) et une formate déshydrogénase aérobie nitrate-inductible (FdhO).

La capacité de biosynthèse de la sélénocystéine de ces organismes a déjà été étudiée, comme nous l'avons vu au chapitre 2. Nous rappellerons juste ici les données expérimentales d'intérêt concernant les partenaires de cette biosynthèse, objets de la présente étude.

### 3.1.1 Les $\text{ARN}_t^{\text{Sec}}$ de *E. coli* et *M. thermoacetica*

Les structures secondaires des  $\text{ARN}_t^{\text{Sec}}$  d'*Escherichia coli* et de *Moorella thermoacetica* sont visibles sur la figure 2.1, page 80, et l'alignement de leur séquence est présent sur le schéma 4.7, page 187, au paragraphe concernant les résultats d'expériences d'empreinte chimique, § 4.3.2.

L' $\text{ARN}_t^{\text{Sec}}$  d'*Escherichia coli* a été longuement étudié, depuis la découverte d'incorporation co-traductionnelle de la sélénocystéine dans les protéines. L'ensemble des données ainsi accumulées a servi de base à notre travail : sa production, ses propriétés et son aminoacylation, notamment, ont été décrites dans de multiples publications.<sup>[II.6-II.10]</sup>

L' $\text{ARN}_t^{\text{Sec}}$  de *Moorella thermoacetica* a été beaucoup moins décrit, mis à part pour souligner sa longueur et son contenu en paire de bases G-C. Il est l'ARN de transfert le plus long connu actuellement, avec 100 nucléotides, et par conséquent également le plus long des  $\text{ARN}_t^{\text{Sec}}$  connus.<sup>[II.11]</sup> Son contenu en paire de bases G-C (24 sur 28 paires de bases appariées) peut s'expliquer par le fait que cet organisme anaérobie est un thermophile, vivant dans un milieu à 58°C. Les paires G-C permettent de stabiliser l' $\text{ARN}_t$  en augmentant sa température de fusion. Ces détails auront leur importance lors de la tentative de production de cet  $\text{ARN}_t$ .

Organisme	Nbre de bases	Masse molaire	$\epsilon$ à 260nm	$\epsilon$ à 280nm
<i>E. coli</i>	95	30568 Da	896100 $\text{M}^{-1}\cdot\text{cm}^{-1}$	403300 $\text{M}^{-1}\cdot\text{cm}^{-1}$
<i>M. thermoacetica</i>	100	32450 Da	927800 $\text{M}^{-1}\cdot\text{cm}^{-1}$	417500 $\text{M}^{-1}\cdot\text{cm}^{-1}$

**Tableau 3.1** – Données caractéristiques des  $\text{ARN}_t^{\text{Sec}}$  étudiés. Les masses molaires et les coefficients d'absorption  $\epsilon$  à 260nm sont des résultats calculatoires. Les coefficients  $\epsilon$  à 280nm ont été obtenus par comparaison de l'absorption à 260 et 280nm, via des mesures spectroscopiques directes.

### 3.1.2 Caractéristiques des sélénocystéine synthèses SelA

Les versions sauvages des sélénocystéine synthase d'*Escherichia coli* et de *Moorella thermoacetica* ont été caractérisées par de nombreuses études, et les données relatives à ces versions sont présentées à la section 2.3, page 90.

Durant ce doctorat, je n'ai pas travaillé sur les versions sauvages de ces enzymes, mais sur leurs versions rallongées d'une étiquette (ou tag) à 6 histidines pour faciliter leur purification par affinité tag-dépendante. (Tous les détails sur le clonage, la production et la purification de ces protéines sera abordé par la suite, § 3.3, page 144.)

Pour des raisons de cohérence, pour toute la suite de ce manuscrit, les sélénocystéines synthèses recombinantes à tag-6His seront nommées respectivement ecSelA et mtSelA pour les protéines d'*Escherichia coli* et de *Moorella thermoacetica*.

Le terme SelA recouvrira ces deux versions, dans les cas où l'une et l'autre de ces protéines sont indifféremment désignées.

Les versions sauvages seront, quant à elles, appelées ecSelA<sub>sauvage</sub> et mtSelA<sub>sauvage</sub>.

En plus des protéines entières taguées, ecSelA et mtSelA, nous avons également produit, lors des tentatives de clonages des gènes correspondants, divers mutants, dont un particulièrement intéressant : mtSelA( $\Delta$ 1-31) qui se caractérise par la délétion des 31 premiers résidus de la séquence de mtSelA.

Les séquences des protéines ecSelA<sub>sauvage</sub> et mtSelA<sub>sauvage</sub> sont représentées, alignées, en annexe A.4, page 232, mais également au § 5.2.2, page 212, où sont présentées les prédictions bio-informatiques des structures secondaires de ces protéines.

Les séquences des protéines ecSelA et mtSelA sont caractérisées par une étiquette de 35 acides aminés supplémentaires, dont 6 histidines terminales, ce qui leur donne les caractéristiques suivantes :

Protéine	Longueur	Masse molaire	pI <sub>(th)</sub>	$\epsilon_{(th)}$ à 280nm
ecSelA <sub>sauvage</sub>	463 aa	50607 Da	6,05	35410 M <sup>-1</sup> ·cm <sup>-1</sup>
ecSelA	498 aa	54327 Da	6,11	35410 M <sup>-1</sup> ·cm <sup>-1</sup>
mtSelA <sub>sauvage</sub>	470 aa	50590 Da	6,75	25900 M <sup>-1</sup> ·cm <sup>-1</sup>
mtSelA	505 aa	54310 Da	6,56	25900 M <sup>-1</sup> ·cm <sup>-1</sup>
mtSelA( $\Delta$ 1-31)	474 aa	50566 Da	6,25	24410 M <sup>-1</sup> ·cm <sup>-1</sup>

**Tableau 3.2** – Données caractéristiques des protéines SelA étudiées. Les masses molaires, les pI<sub>(th)</sub> et les coefficients d'absorption  $\epsilon_{(th)}$  à 280nm sont des résultats calculatoires et ne tiennent pas compte de la présence de PLP — dont la masse molaire est de 246 Da.

## 3.2 Obtention des $\text{ARN}_t^{\text{Sec}}$

Pour les besoins des différentes expériences, nous avons produit des quantités importantes d' $\text{ARN}_t^{\text{Sec}}$  d'*E. coli*, et également dans une moindre mesure de *M. thermoacetica*, dans le but de les utiliser tels quels pour des études d'interactions, et chargés en sérine, *ie.* sous forme de séryl- $\text{ARN}_t^{\text{Sec}}$ , pour la formation du complexe avec SelA.

Chronologiquement, la production en quantité de l' $\text{ARN}_t^{\text{Sec}}$  d'*Escherichia coli* précéda celle de l' $\text{ARN}_t^{\text{Sec}}$  de *Moorella thermoacetica* de près de deux ans, pour des raisons d'antécédence de l'usage de celui d'*E. coli* au laboratoire, mais aussi pour des raisons techniques sur lesquelles nous reviendrons plus loin.

La production des séryl- $\text{ARN}_t^{\text{Sec}}$  fut, lors de mon doctorat, à la fois une tâche de routine consommatrice de temps et un ensemble de manipulations parmi les plus sensibles à maîtriser. En effet, lorsque aucune mise au point n'est nécessaire, l'ensemble des opérations nécessaires à la production d'un lot d' $\text{ARN}_t^{\text{Sec}}$  purifié nécessite 4 jours de travail. La charge du lot en question demande une mise au point préliminaire par cinétique et occasionne 2 jours de manipulations supplémentaires.

À côté de cela, les séryl- $\text{ARN}_t$  sont connus pour être les  $\text{ARN}_t$  les plus sensibles et les plus rapides à la déacylation.<sup>[II.12]</sup> Nous avons donc choisi de produire les séryl- $\text{ARN}_t^{\text{Sec}}$  par lot minimum pour emploi quasi-immédiat, et de répéter la production aussi souvent que nécessaire. Cette précaution est également l'assurance d'une perte moindre de matériel en cas de dégradation de la production.

Cette stratégie a aussi permis une amélioration successive de la technique, et une moindre conséquence des problèmes techniques qui peuvent influencer le rendement global de la production, celui-ci ayant balayé, durant mon doctorat, la plage de  $26\mu\text{g}$  à  $342\mu\text{g}$  d' $\text{ARN}_t^{\text{Sec}}$  produits par mL de transcription.

Les différentes étapes de la production d'un lot d' $\text{ARN}_t^{\text{Sec}}$  sont exposées ci-dessous, et leurs protocoles détaillés sont présentés en Annexe B.2, page 247.

### 3.2.1 Au départ, *selC*...

Il existe plusieurs voies de production des  $\text{ARN}_t$  :

- L'extraction, à partir de cultures bactériennes, des quantités naturelles d' $\text{ARN}_t$  présentes dans les cellules, puis purification de l'espèce d' $\text{ARN}_t$  désirée.
- La production par culture d'*Escherichia coli* recombinante, possédant un plasmide pour la surexpression d'une espèce d' $\text{ARN}_t$ . La population de cet  $\text{ARN}_t$  est alors majoritaire : la purification s'en trouve facilitée et les quantités obtenues beaucoup plus conséquentes.
- La synthèse chimique sur support solide, qui permet également d'obtenir des  $\text{ARN}$ , mais dont le rendement, au-delà de 30 nucléotides, diminue sensiblement. Ce n'est donc pas



une méthode de choix pour produire les ARN<sub>t</sub>, la longueur de ceux-ci conduisant à un rendement médiocre.

- La synthèse enzymatique *in vitro*. Cette méthode est celle que nous avons utilisée le plus pour tous les ARN<sub>t</sub><sup>Sec</sup> qui ont été utilisés lors des expériences biochimiques ou de cristallogénèse. Cette méthode consiste en deux étapes, l'obtention d'une matrice d'ADN par PCR, puis la transcription enzymatique de cette matrice en ARN<sub>t</sub>.

### Production de la matrice d'ADN de transcription

Lors de la transcription enzymatique, la polymérase T7 utilise comme patron de synthèse une matrice d'ADN linéaire contenant le gène de l'ARN à synthétiser. Cette matrice doit comporter, en plus du gène *selC*, la région promotrice de la transcription, constituée de 17 nucléotides, en amont du gène.

Pour la production de cette matrice, nous disposons du plasmide contenant le gène *selC* d'*E. coli*, obtenu par un clonage classique.

Ce plasmide est transmis par électroporation à une souche d'*Escherichia coli* XL1 Bleu, qui se charge de l'amplification de ce dernier. Les plasmides obtenus par culture de ces bactéries transformées sont récupérés par lyse et purification grâce au kit commercial de réactifs prévus à cet effet.

Le plasmide contenant le gène *selC* d'*Escherichia coli* est utilisé comme matériel de départ de la PCR aboutissant à la matrice d'ADN linéaire qui nous intéresse. Pour ce faire, les amorces de PCR utilisées sont :

- le complémentaire de la séquence 5'-terminale de l'ARN<sub>t</sub><sup>Sec</sup> sur une vingtaine de nucléotides, 5'[TGGCGGAAGATCACAGGAGTCG]<sup>3'</sup> ;
- et la séquence 3'-terminale de l'ARN<sub>t</sub><sup>Sec</sup> sur autant de nucléotides, précédée du promoteur de transcription T7, 5'[TAATACGACTCACTATAGGAAGATCGTTCGTCTCCGGTGAGG]<sup>3'</sup>.

Lors de la PCR, les vingt nucléotides correspondant à l'ARN<sub>t</sub><sup>Sec</sup> s'apparient au gène *selC* du plasmide, qui devient le patron de l'élongation de l'amorce. Lors du deuxième cycle de PCR, ce nouveau brin d'ADN contenant le gène *selC* dans son intégralité et le promoteur T7 sert de patron pour une extension de l'amorce complémentaire.

Cette PCR de 25 cycles, effectuée par l'ADN-polymérase Taq, aboutit ainsi à une population ultra-majoritaire de la matrice linéaire double brin attendue, contenant le promoteur T7 suivie du gène *selC*.

Les amorces utilisées sont obtenues par synthèse chimique sur support solide, production qui est couramment externalisée, des entreprises spécialisées proposant ce service. Il est possible d'obtenir par ce biais des amorces contenant des modifications chimiques. Une modification est particulièrement utile dans notre cas : des nucléotides **ribonucléiques** méthylés sur leurs hydroxyles 2' peuvent être positionnés en fin de séquence transcrite, c-à-d. à l'extrémité 5' de l'amorce complémentaire à la séquence de l'ARN<sub>t</sub>. Cette modification ne perturbe pas la PCR, mais permet d'arrêter à leur niveau le fonctionnement

de la T7 polymérase, qui a tendance, lors de la transcription, à ajouter des nucléotides surnuméraires à l' $ARN_t$  synthétisé, le rendant ainsi inutilisable. La quantité et le positionnement des nucléotides ainsi modifiés dans la séquence de l'amorce dépendent du gène à transcrire, et font l'objet d'une étude lors des premières transcriptions.

Dans notre cas, une unique guanine modifiée en pénultième position suffit pour l'obtention d'une proportion importante de terminaison de la transcription à la position attendue.

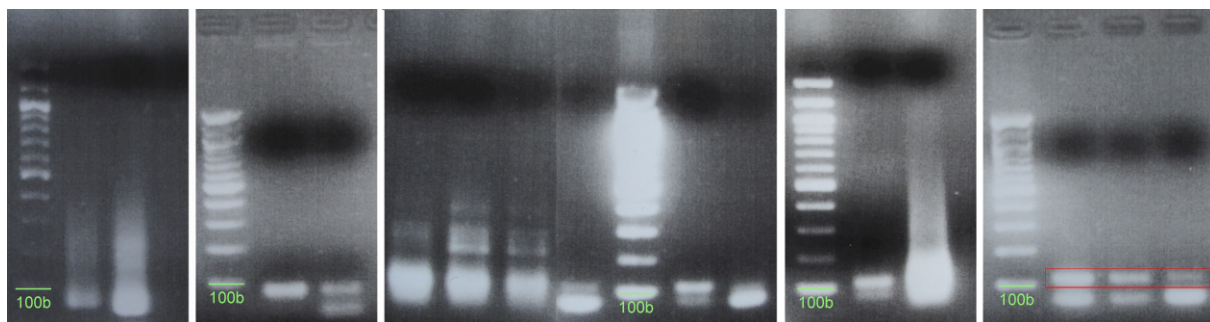
La matrice ADN obtenue par PCR est purifiée par extraction de la phase aqueuse de réaction par un même volume d'un mélange phénol/chloroforme/alcool isoamylique (25:24:1), puis par un volume de chloroforme. L'ADN est ensuite précipité par ajout d'éthanol à la solution (2,5 volumes). La centrifugation du précipité permet sa récupération et son séchage, pour une redissolution en une solution concentrée, tamponnée à pH 8,0. L'échantillon obtenu peut être stocké à  $-20^{\circ}C$  pour une utilisation ultérieure en transcription ou pour une nouvelle PCR.

En théorie, il serait possible d'effectuer indéfiniment des PCR de la matrice linéaire d'ADN obtenue, avec les mêmes amorces. Mais en pratique il est conseillé de partir le plus souvent possible de plasmides ou de produits de première PCR, l'ADN-polymérase Taq n'étant pas une enzyme à fidélité excellente.

Il peut arriver, pour le plus grand malheur du manipulateur, que certaines séquences d'ADN ne soient pas facilement amplifiables.

Ce fut le cas pour la matrice de transcription de l' $ARN_t^{Sec}$  de *Moorella thermoacetica*. Des exemples de gels d'agarose montrant les mauvais résultats de PCR obtenus, sont présentés ci-après.

On peut y voir que la PCR de *selC* de *M. thermoacetica* aboutie à : des échantillons contenant une population majoritaire de matrice de taille inférieure à la taille attendue, des échantillons à multiples populations et des échantillons caractérisés par une traînée sur le gel d'agarose, indiquant la non-spécificité de la population des brins d'ADN obtenus.



**Figure 3.1** – PCR d'obtention de la matrice contenant *selC* de *M. thermoacetica*. Le premier échelon de l'échelle d'échantillons de référence correspond à une taille de 100 bases, la matrice d'ADN obtenue devrait avoir 117 bases. Après une première amélioration du protocole, les résultats obtenus (à droite, entre autres) montrent la présence de deux populations de brins d'ADN de tailles différentes, dont la matrice d'intérêt.

Une des hypothèses concernant le dysfonctionnement de ces PCR est liée à l'originalité de *selC* de *M. thermoacetica* : comme indiqué précédemment, le gène de cet  $ARN_t^{Sec}$  est caractérisé par sa longueur et par sa contenance en paires de bases G-C. En particulier, le début de sa séquence, en 5', est composé d'une enfilade de 5 guanines. Or il est connu que les poly-G ont tendance à construire, dans certaines conditions, des montages à plusieurs brins, à très longue durée de vie et à température de fusion élevée.

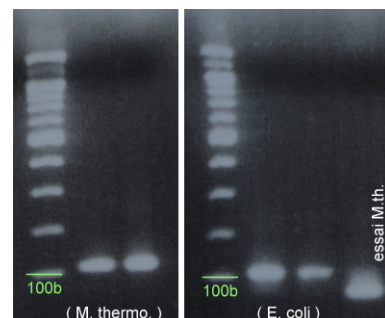
J'ai eu la chance de disposer d'une des meilleures sources de renseignements sur la formation de ces constructions, les quadruplexes de poly-G étant un sujet d'étude de Jean-Louis Leroy,<sup>[II.13]</sup> membre de notre laboratoire. J'ai ainsi appris que selon les conditions de salinité et de concentration, la chaîne poly-G constitutive de l'amorce ou de la matrice d'ADN produite au cours de la PCR peuvent s'associer en construction tétramérique stable. Les conséquences sur la PCR de ce comportement peuvent être catastrophiques : arrêt de la réaction de polymérisation enzymatique de l'ADN en milieu de séquence, séquestration des amorces et des matrices linéaires déjà répliquées, blocage du processus de réplication et d'amplification.

D'autres hypothèses ont été émises pour expliquer les mauvais résultats de PCR observés, dont une multiplicité des sites de fixation de l'amorce sur le plasmide, une mauvaise qualité des amorces utilisées, ou une longueur trop importante de l'amorce 5'-terminale qui formerait un double brin à température de fusion trop élevée pour un bon fonctionnement de la PCR.

Après étude des multiples paramètres influant sur les résultats de PCR — température de l'étape d'hybridation, type de sels contenus dans le tampon de PCR (KCl, LiCl ou  $(NH_4)_2SO_4$ ), longueur de l'amorce 5' terminale utilisée et mutation de la région poly-G — les meilleurs résultats furent obtenus dans les conditions strictes suivantes :

- la source de PCR de *selC* ne peut être que le plasmide purifié ;
- l'amorce 5'-terminale doit être de longueur réduite : 19 bases pour l'appariement avec le gène *selC*, en plus des 17 bases du promoteur T7 (les mutations de séquence poly-G ne sont pas nécessaires) ;
- la salinité du tampon doit être fournie sous forme de LiCl, ou de KCl avec des résultats légèrement moins bons ;
- la température d'hybridation doit être augmentée et passer de 55°C à 58°C, voire à 60°C, avec pour conséquence une meilleure résolution de la population de brins d'ADN obtenus ;
- et la concentration des amorces de PCR dans le milieu de réaction doit être divisée par 2, et cela par dilution de la solution stock de l'amorce 5'-terminale de 100µM à 20µM et correction du volume employé en conséquence.

Les conséquences de l'utilisation d'un tel protocole sont un résultat de PCR homogène, d'une seule population de la taille attendue.



**Figure 3.2** – Comparaison entre une PCR réussie de *selC* de *M. thermoacetica* (à gauche), et une PCR de *selC* d'*E. coli* et un essai avec *M. thermoacetica* (à droite).

Nous avons donc au final réussi à obtenir les matrices de transcription des  $ARN_t^{Sec}$  d'*E. coli* et de *M. thermoacetica*. La réaction de transcription qui s'en suit permet d'apprécier la qualité des lots d'ADN obtenus, et ainsi de valider les protocoles mis en place.

### 3.2.2 La transcription

La réaction de transcription est effectuée par la polymérase T7 à 37°C. Il s'agit d'une enzyme disponible commercialement. Cependant, les quantités d'enzymes nécessaires pour une production d'ARN en vue d'une étude structurale sont relativement importantes. Cette enzyme est donc produite et purifiée, selon les besoins, au laboratoire.

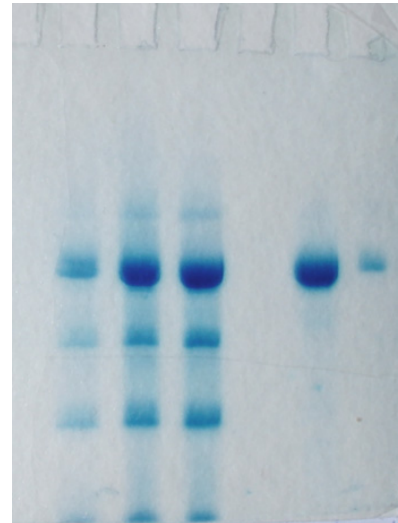
La réaction de transcription *in vitro* se déroule en trois étapes. La première étape est une étape d'initiation, pendant laquelle la polymérase T7 se fixe sur le double brin de la matrice ADN, au niveau de la région promotrice, le dissocie et commence la transcription. Puis se déroule une étape de transition au cours de laquelle la synthèse peut s'arrêter et donner lieu à des transcrits courts dits abortifs. Enfin la transcription devient processive au-delà de douze nucléotides : c'est l'élongation. En fin de transcription, il arrive que la polymérase T7 ajoute des nucléotides surnuméraires par rapport à la séquence à synthétiser, mais ce phénomène est inhibé, avec une bonne efficacité, par la présence de nucléotides modifiés sur la matrice d'ADN, comme décrit précédemment.

Là aussi, le produit de réaction est purifié par une extraction de la phase aqueuse de transcription au phénol puis au chloroforme. Cette extraction est suivie d'une précipitation dans une solution d'éthanol à 70%. La centrifugation du précipité à 20000g permet sa récupération et son séchage, pour une redissolution en une solution concentrée.

La transcription est une méthode enzymatique, qui doit être améliorée en fonction de la séquence à transcrire. Un travail d'optimisation a donc été effectué au cours de mon doctorat, avec comme résultat notable l'augmentation d'un facteur 14 du rendement de la transcription pour l' $ARN_t^{Sec}$  d'*E. coli*, comme indiqué en introduction.

Cette augmentation de rendement résulte :

- d'une purification minutieuse de la matrice ADN obtenue dans de bonnes conditions de PCR ;
- de l'ajustement du ratio entre la concentration de ribonucléotides-triphosphates et de magnésium ( $MgCl_2$ ) présents dans le milieu réactionnel ;
- et de l'ajout à la réaction de pyrophosphatase, qui, par dégradation du pyrophosphate en phosphate, permet le déplacement de la réaction vers un avancement plus important.

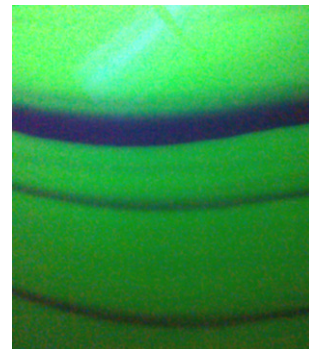


**Figure 3.3** – Gel de polyacrylamide montrant des résultats de transcription. De gauche à droite, trois différents protocoles de transcription, l' $ARN_t^{Sec}$  après purification, ainsi qu'un repère pour le gel.

En plus d'un meilleur rendement, l'amélioration de la réaction a pour conséquences directes la diminution des transcrits abortifs et l'augmentation de la vitesse de réaction.

### Purification de l'ARN transcrit

La seule méthode de purification permettant de séparer des ARN n'ayant que quelques nucléotides de plus ou de moins, est la purification sur gel dénaturant de polyacrylamide / urée à 7M, que nous coulons et polymérisons au laboratoire. Comme la quantité d' $ARN_t^{Sec}$  que nous synthétisons est importante (de l'ordre de 2 mg par lot), la taille des gels utilisés pour la purification des transcriptions est conséquente : 50cm x 30cm x 3mm, soit quasiment 500 mL de mélange à polymériser par gel.



Après électrophorèse, démoulage et découpage des bandes d'ARN d'intérêt, cet ARN est électro-élué hors du polyacrylamide dans des bains compartimentés, une cellule de séquestration délimitée par des membranes adéquates permettant la récupération de l' $ARN_t^{Sec}$ .

**Figure 3.4** – Grand gel d'acrylamide de purification de l' $ARN_t^{Sec}$ , démoulé et observé sous U.V.

Après précipitation par les mêmes méthodes que précédemment, l' $ARN_t^{Sec}$  est stocké sous forme de solutions concentrées.

### 3.2.3 La charge en sérine

Les  $ARN_t^{Sec}$  d'*E. coli* et de *M. thermoacetica* peuvent être utilisés tels quels, pour des expériences d'interactions simples entre Sela et l'ARN non-acylé. Cependant, pour la formation du complexe stable qui nous intéresse, l'estérification d'une sérine à l'extrémité 3' de l' $ARN_t^{Sec}$  est nécessaire.

La méthode de choix pour effectuer cette réaction est la voie bio-mimétique, utilisant l'enzyme naturelle de cette réaction, la Ser- $ARN_t$  synthétase d'*Escherichia coli*. Celle-ci devrait charger en sérine de façon quasiment indifférente les  $ARN_t^{Sec}$  d'*E. coli* et de *M. thermoacetica*, avec tout de même une activité assez faible. (voir § 2.1.1, page 83)

### Production de la séryl- $ARN_t^{Sec}$ synthétase

À mon arrivée, le laboratoire disposait déjà du plasmide contenant une version tag-6His de la SerRS d'*Escherichia coli*. Un travail préliminaire à la charge des  $ARN_t^{Sec}$  consiste à transformer des souches d'*E. coli* BL21 DE3 thermo-compétentes avec ce plasmide, à exprimer et à purifier la protéine.

Cette protéine est très bien sur-exprimée en bactéries recombinantes. La quantité de protéine produite est telle qu'une unique purification par colonne d'affinité au nickel suffit pour obtenir une protéine d'une pureté adéquate. Il est toutefois nécessaire de vérifier l'absence d'activité RNase de l'échantillon ainsi purifié, comme cela est fait pour les échantillons de SelA (voir § 3.3.3.2, page 146).

### Cinétique et profil de charge

L'estérification enzymatique des  $ARN_t^{Sec}$  par la SerRS s'effectue dans un tampon particulier appelé "Pure System Buffer" à cause de sa provenance du système purifié de traduction protéique *in vitro*. Ce tampon est censé reproduire entre autres les conditions de salinité et de pH de la cellule (pH = 7,3).

Malgré ces conditions bio-mimétiques, la catalyse de l'estérification de la sérine sur l' $ARN_t^{Sec}$  par la SerRS n'est pas optimum. Le cas est assez complexe :

- La réaction catalysée par la SerRS est réversible : il est nécessaire d'effectuer une cinétique de la réaction sur un petit échantillon pour connaître la durée au bout de laquelle l'avancement est optimal. Cette durée de réaction est fonction de différents facteurs comme le lot et la concentration de l'enzyme et des  $ARN_t$  utilisés, ou la concentration d'ATP et de sérine présents lors de la réaction.

- Le suivi de la réaction ne peut se faire que selon une seule méthode, destructrice et nécessitant l'emploi de sérine<sup>14</sup>C. Lors de cette réaction d'estérification à petite échelle avec de la sérine<sup>14</sup>C, des prélèvements d'un volume précis de milieu réactionnel sont effectués à des temps donnés. Pour stopper la réaction, le volume prélevé est déposé sur du papier Whatman®, qui est immédiatement plongé dans un bain d'acide trichloroacétique, permettant la précipitation des  $ARN_t^{Sec}$  sur le papier et le lavage de la sérine<sup>14</sup>C résiduelle.

La quantité de radioactivité présente sur le papier, fixée à l'ARN, est comptée par scintimétrie. Le rapport de la quantité de radioactivité mesurée sur la quantité de radioactivité attendue (selon la concentration d' $ARN_t^{Sec}$ ) donne l'avancement de la réaction.

- Une fois les paramètres optimaux de réaction cernés et le meilleur temps d'avancement connu, la réaction se fait sur le lot entier, en aveugle. La purification du séryl- $ARN_t^{Sec}$  fonctionne sur le même principe que celle de l' $ARN_t^{Sec}$  classique. Lors de la précipitation à l'éthanol à 70%, l'ATP et l'AMP de la réaction co-précipitent avec l' $ARN_t^{Sec}$  : il est important de connaître la quantité d' $ARN_t^{Sec}$  mis en réaction car celle-ci ne pourra plus être mesurée par la suite, via les méthodes spectroscopiques habituelles. Il est surtout impossible de connaître la proportion d' $ARN_t^{Sec}$  chargé en sérine par rapport à l' $ARN_t^{Sec}$  déacylé ou non chargé : la cinétique effectuée avec de la sérine<sup>14</sup>C procure la seule estimation du rendement de la réaction "froide".

Parmi les résultats obtenus, la figure 3.5, page 143 reprend ceux de certaines expériences de cinétique "chaudes", illustrant ainsi le fonctionnement de la réaction.

Comme montré en (C.), la cinétique est effectuée par mesure de l'avancement de la réaction à des temps de 5, 10, 15 et 20 minutes. Toujours dans le cas illustré en (C.), l'activité spécifique de la sérine<sup>14</sup>C et le volume de réaction utilisé pour la mesure à chaque temps donné, permettent de calculer que la charge théorique de 100% produit une radioactivité de 4000 CPM pour 1,0  $\mu$ M d'ARN<sub>t</sub><sup>Sec</sup>, de 2800 CPM pour 0,7  $\mu$ M d'ARN<sub>t</sub><sup>Sec</sup> et de 1600 CPM pour 0,4  $\mu$ M d'ARN<sub>t</sub><sup>Sec</sup>.

En (A.) figure le pourcentage de séryl-ARN<sub>t</sub><sup>Sec</sup> obtenu par rapport à la quantité attendue, en fonction du lot et de la concentration de l'ARN<sub>t</sub><sup>Sec</sup> mis en réaction. Cet avancement de réaction est obtenu pour la durée optimale de réaction, qui est généralement de 15 minutes. Alors que le lot d'ARN<sub>t</sub><sup>Sec</sup> n°3 sera acylé dans une réaction avec 0,7  $\mu$ M de matériel, le lot n°2 bénéficiera d'une acylation à la concentration de 1,0  $\mu$ M. Le lot n°4, présentant visiblement une anomalie, sera re-purifié et caractérisé par des nouvelles mesures.

En (B.) sont présentées les mesures effectuées après une purification de SerRS : il est notablement étonnant que la réaction atteigne un avancement plus élevé à une concentration moindre de l'enzyme, ce qui sera pris en compte pour les réactions de charge effectuées avec ce lot d'enzyme.

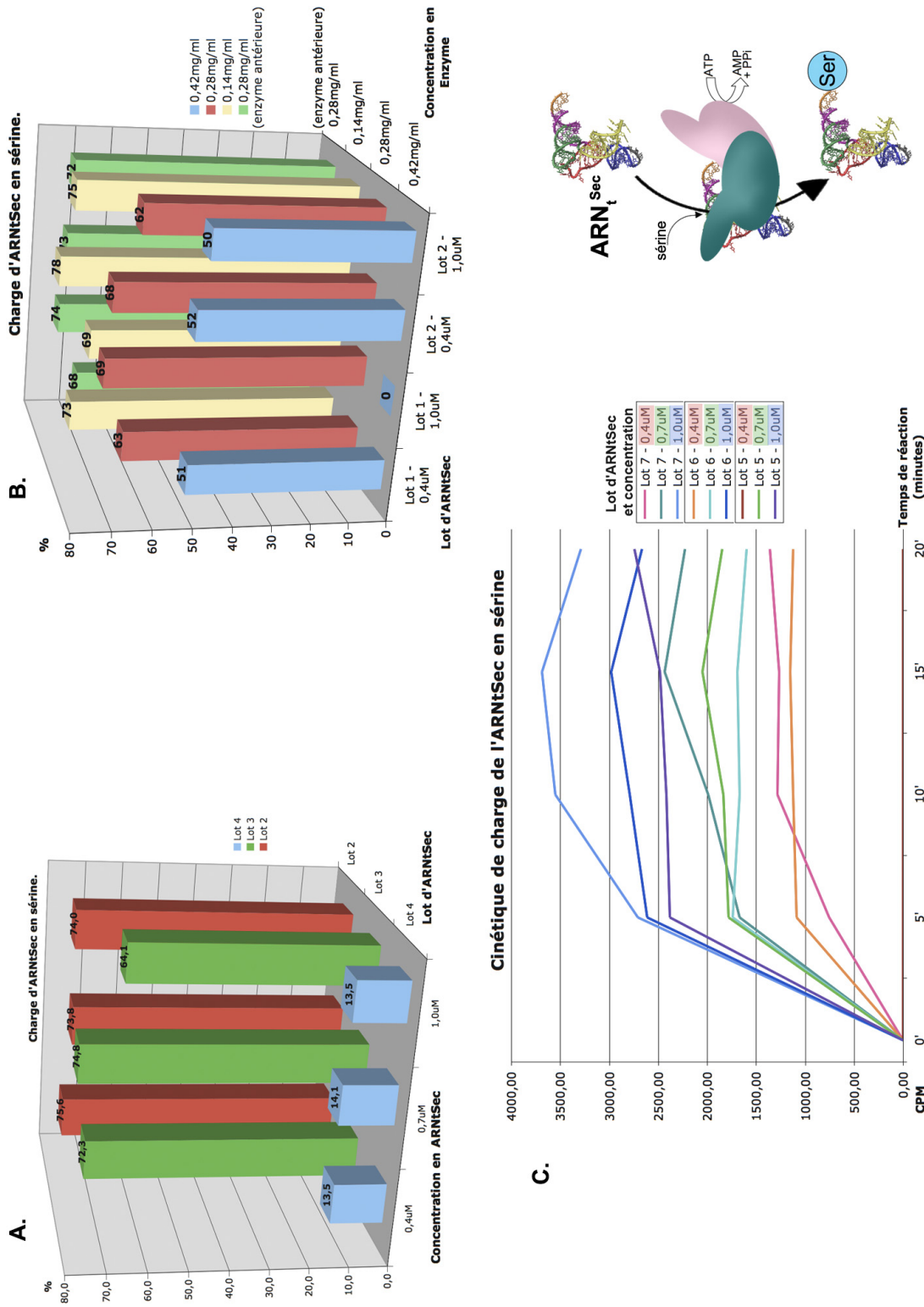
On constate ainsi qu'une charge moyenne atteint seulement 75% d'avancement. À l'heure actuelle, la méthode de charge employée est légèrement différente et permet d'obtenir facilement, lors de la maîtrise des différents paramètres sensibles énumérés précédemment, 95% de charge observée. L'amélioration de la réaction passe par l'ajout de pyrophosphatase au milieu réactionnel, qui permet, comme lors de la transcription, une hydrolyse du pyrophosphate en deux phosphates, et tire ainsi l'équilibre de la réaction vers la formation du séryl-ARN<sub>t</sub><sup>Sec</sup>. Cette méthode fut employée, par exemple, lors de la charge du lot n°7 dont la cinétique est visible en (C.) de la figure 3.5, page suivante.

### 3.2.4 ... au final, le séryl-ARN<sub>t</sub><sup>Sec</sup>

Les différents lots de séryl-ARN<sub>t</sub><sup>Sec</sup> obtenus sont stockés à -20°C dans un tampon d'acétate de sodium à un pH entre 5,0 et 5,5. Cette précaution est importante pour la stabilité dans le temps de ces ARN<sub>t</sub><sup>Sec</sup>, car les ARN<sub>t</sub> se déacylent spontanément à pH basique, voire neutre. Nous l'avons rappelé précédemment, le séryl-ARN<sub>t</sub> est le plus labile des ARN<sub>t</sub> : sa conservation en tampon acide est donc obligatoire.

Des mesures de la cinétique de déacylations, en fonction de la nature et du pH du tampon dans lequel était dissous le Ser-ARN<sub>t</sub><sup>Sec</sup>, ont été effectuées pour confirmer l'importance de ces précautions et seront discutées au § 4.2, page 179.

Nous verrons par la suite que la nécessité d'utiliser les séryl-ARN<sub>t</sub><sup>Sec</sup> en milieu acide peut poser problème lors de la formation du complexe avec la sélénocystéine synthase SelA.



**Figure 3.5** – Données de cinétique et d'avancement de charge à la sérine de lots d'ARNt<sup>Sec</sup> d'*E. coli* par la SerRS. (A.) Avancement de la charge en fonction du lot d'ARNt<sup>Sec</sup> et de sa concentration lors de la réaction. (B.) Avancement de la charge en fonction du lot et de la concentration de l'enzyme SerRS. (C.) Cinétique de charge de trois lots d'ARNt<sup>Sec</sup> d'*E. coli* à des concentrations différentes.



## 3.3 Obtention des sélénocystéine synthases

### 3.3.1 Clonage et vecteur d'expression

Comme indiqué précédemment, les protéines SelA d'*Escherichia coli* et de *Moorella thermoacetica* utilisées pour notre étude sont des versions rallongées d'une étiquette à 6 histidines terminales.

C'est par un clonage et une ligation classiques que les versions sauvages des gènes *selA* de ces deux organismes sont introduites dans des plasmides pET29b, permettant une expression des protéines grâce à un promoteur T7, sous le contrôle d'un opérateur *lac*. La carte de ce plasmide est visible en annexe B.1.2, page 246. La construction employée permet de positionner les 6 histidines en position C-terminale de l'enzyme, après un raccord de 29 acides aminés.

Ce long raccord permet aux 6 histidines d'être facilement accessibles, et donc de faciliter l'adsorption de la protéine recombinante sur la colonne d'affinité lors de sa purification. De plus, un site de coupure à la thrombine est présent sur ce raccord, ce qui peut permettre la suppression de l'étiquette après la première étape de purification, si celle-ci est gênante pour l'interaction entre SelA et l'ARN<sub>t</sub><sup>Sec</sup>.

Une fois la construction désirée obtenue, des souches d'*E. coli* XL1-Blue sont transformées par électroporation avec le plasmide, puis cultivées sur boîte de Pétri. Une sélection par résistance à l'antibiotique kanamycine, induite par le plasmide, permet l'isolement des clones effectivement transformés, qui sont ensuite cultivés en milieu de culture classique LB (milieu de Luria-Bertani).

La culture bactérienne permet ainsi l'amplification du plasmide, qui sera récolté après lyse cellulaire, par les techniques habituelles déjà évoquées dans le cas du gène *selC* codant pour l'ARN<sub>t</sub><sup>Sec</sup>.

### 3.3.2 Production et récolte

Les protéines ecSelA, mtSelA et mtSelA( $\Delta$ 1-31) sont produites de manières usuelles, sans poser de difficultés particulières.

Des souches thermo-compétentes d'*E. coli* BL21 DE3 sont transformées avec les plasmides respectifs, puis sont cultivées sur boîte de Pétri. Des clones isolés sont sélectionnés pour devenir les sources ultérieures de cultures en milieu liquide de grande dimension.

L'opérateur *lac*, présent sur le plasmide entre le promoteur T7 et le gène de la protéine, permet une répression de l'expression du gène *selA* pendant la culture bactérienne jusqu'à sa phase de croissance exponentielle. Lorsque cette phase est atteinte, l'expression du gène est induite par l'ajout dans le milieu d'IPTG (Isopropyl  $\beta$ -D-1-thiogalactopyranoside), que fixe le répresseur *lac* au même titre que le lactose, mais qui n'est lui pas dégradé par le

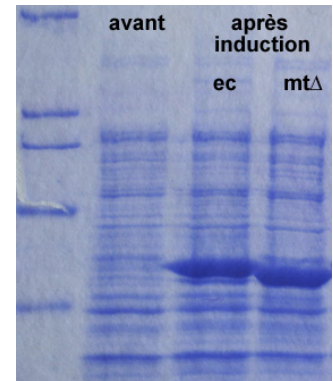
métabolisme de la cellule. Le répresseur libère ainsi l'opérateur *lac*, ce qui provoque la transcription du gène dans la cellule hôte et la production de la protéine d'intérêt.

Après 4h d'induction de la production de la protéine, les bactéries sont culottées par centrifugation, puis lysées par ultrasons ou par l'utilisation d'un "casseur de cellules à pression constante" de Constant Systems®, dans le tampon de lyse également utilisé lors de la première étape de purification. Ce tampon est très salin (1M de  $\text{NH}_4\text{Cl}$ ) pour favoriser la lyse des cellules, et contient un minimum d'imidazole, composé qui joue le rôle de solubilisant de la protéine sur-exprimée.

En effet, celle-ci est particulièrement lipophile et a tendance à s'adsorber sur les débris cellulaires, en plus d'être produite par la cellule en partie sous forme de corps d'inclusion insolubles. Même si un simple lavage des débris cellulaires avec un tampon contenant des détergents permet une meilleure solubilisation de la protéine dans le surnageant, cette technique n'est pas utilisée, pour préserver l'homogénéité et la pureté de l'échantillon obtenu, les détergents pouvant être difficiles à extraire du tampon lors de la purification.

Ce problème de lipophilie et de tendance à l'agrégation nécessite également, lors du culottage des débris cellulaires par centrifugation, que la vitesse de centrifugation ne dépasse pas 9000g, pour conserver une majorité de protéines sur-exprimées dans le surnageant de l'échantillon.

Ce surnageant est immédiatement filtré, et purifié sur colonne d'affinité au tag histidine.



**Figure 3.6** – Gel SDS PAGE du lysat cellulaire avant et après induction de l'expression de *ecSelA* et de *mtSelA*( $\Delta 1-31$ ).

### 3.3.3 Purification des constructions protéiques

#### 3.3.3.1 Méthode initiale

La méthode de purification employée au laboratoire pour la protéine *ecSelA* est basée sur l'efficacité de la colonne d'affinité au tag-6His. Cette colonne permet en effet de retenir les protéines munies du tag en question, et de les laver des protéines cellulaires également présentes dans le surnageant du lysat. La rétention des protéines possédant le tag-6His est basée sur la complexation du nickel ( $\text{Ni}^{2+}$ ) par les histidines successives présentes à la fin C-terminale des constructions protéiques sur-exprimées. Les colonnes d'affinité au tag-6His sont pour cela chargées d'ions nickel fixés aux billes de la colonne par complexation avec des chaînes fonctionnalisées adéquates.

La protéine d'intérêt est ensuite décrochée de la colonne par une augmentation, dans le tampon d'éluion, de la concentration en imidazole, qui est un compétiteur de complexation du nickel.

La pureté de l'échantillon protéique obtenu n'est tout de même pas suffisante pour les applications auxquelles nous le destinons : la cristallogénèse demande l'utilisation de matériel homogène et de pureté importante, la complexation de la protéine avec de l' $\text{ARN}_t^{\text{Sec}}$  nécessite l'absence de contaminants protéiques ayant une activité RNase.

La première tentative de purification complémentaire que nous avons effectuée, est l'utilisation d'une colonne CM (carboxyméthyle), échangeuse de cation. Il a en effet été montré<sup>[II.14]</sup> que la protéine sauvage, ecSelA<sub>sauvage</sub>, est chargée positivement en surface, ce qui est en accord avec sa fonction d'interaction avec l'ARN<sub>t</sub><sup>Sec</sup>, dont le squelette phosphate est chargé négativement. Une colonne échangeuse de cation peut donc permettre de la séparer des autres protéines contaminantes aux propriétés différentes.

Pour effectuer cette étape de purification, la protéine est tout d'abord dialysée contre un tampon phosphate, en accord avec la littérature.<sup>[II.9, II.10, II.14]</sup>

Malgré une différence mineure entre ecSelA<sub>sauvage</sub> et ecSelA, consistant en l'étiquette contenant le tag-6His, le comportement de cette dernière vis-à-vis du changement de tampon est radicalement différent (voir la figure 3.7). Les facteurs provoquant la précipitation de la protéine ecSelA nous étant inconnus, des essais de tampons aux caractéristiques différentes ont été effectués pour connaître les conditions de stabilité de la protéine. (voir § 3.3.3.3, page ci-contre)

Le surnageant de dialyse contient tout de même une quantité suffisante de protéines pour effectuer un essai de purification sur la colonne CM.

La fraction d'élution d'ecSelA fut testée pour son activité RNase, après concentration.



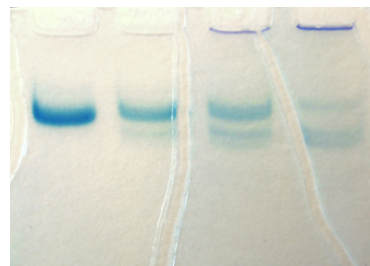
**Figure 3.7** – Précipitation de ecSelA lors d'une dialyse contre un tampon phosphate.

### 3.3.3.2 Activité RNase des échantillons

Pour tester l'activité RNase des échantillons de SelA, un protocole simple a été mis en place : une petite quantité d'ARN<sub>t</sub><sup>Sec</sup> (2µg) est "sacrifiée" lors d'une incubation à 37°C pendant une nuit avec des quantités croissantes de protéines du lot à tester.

Dans l'exemple illustré par la figure 3.8, le lot de ecSelA testé est celui purifié lors du premier essai de purification précédent. Il est présent, dans les échantillons mis à incuber, à hauteur de 0µg, 3µg, 6µg et 15µg, de gauche à droite sur le gel présenté ci-contre.

Une activité RNase est observable : l'ARN<sub>t</sub><sup>Sec</sup> se dégrade proportionnellement à la quantité de protéine mise en sa présence. Le protocole de purification mis en place n'est donc pas suffisamment efficace, même si tous les lots de protéines purifiées grâce à ce premier protocole ne se révèlent pas contaminés.



**Figure 3.8** – Exemple d'activité RNase d'un échantillon de SelA d'*E. coli* sur de l'ARN<sub>t</sub><sup>Sec</sup>.

J'ai décidé à la suite de cette observation que tous les échantillons de protéines purifiées devaient être soumis à un test similaire, pour vérifier leur innocuité vis-à-vis de l'ARN, que seuls les échantillons sûrs seraient employés dans les essais de cristallogenèse, et que de nouvelles méthodes de purification devaient être mises au point.

### 3.3.3.3 Stabilité, propriétés des protéines

Pour pouvoir efficacement mettre au point un protocole de purification d'une protéine, il est nécessaire de connaître sa stabilité vis-à-vis des tampons susceptibles d'être employés.

Nous avons pu remarquer que la protéine ecSelA précipitait lors du passage du tampon de purification sur colonne d'affinité (Tampon A : HEPES 50mM à pH 7,6 ; Imidazole ~300mM ; KCl 100mM ; MgCl<sub>2</sub> 10mM ; DTT 1mM) à un tampon phosphate aux propriétés pourtant presque identiques (Tampon B : phosphate 50mM à pH 7,6 ; Imidazole 300mM ; KCl 50mM ; DTT 1mM).

J'ai donc testé la stabilité dans le temps, à 4°C, d'échantillons de ecSelA identiques, issus de la purification sur colonne d'affinité, vis-à-vis de la présence de phosphate, du pH et de la concentration en imidazole du tampon dans lequel ils étaient placés. La solution témoin utilisée est un de ces échantillons, conservé dans le tampon A, dans lequel la stabilité de ecSelA a pu être observée. La mesure de l'absorption U.V. à 280nm de la solution de départ et du surnageant des solutions centrifugées après 7 jours à 4°C, permet d'avoir une indication qu'en à l'influence des paramètres testés sur la stabilité de la protéine.

Les conclusions de ces tests sont les suivantes :

- La présence de phosphate dans le tampon, même en faible quantité (5mM), provoque la précipitation d'une partie de la protéine ecSelA.
- La protéine est stable dans des solutions contenant de 500mM à 100mM d'imidazole ; à partir de 50mM en imidazole la protéine tend à précipiter, et elle n'est absolument pas stable lorsque le tampon ne contient que des traces d'imidazole. La concentration minimum en imidazole a été déterminée à 80mM.
- Le pH a une forte influence sur la stabilité de la protéine. Pour rappel, le pI théorique de ecSelA est de 6,11. Pourtant les mesures effectuées montrent que ecSelA est stable à un pH compris entre 8,5 et 7,2. La protéine commence à être déstabilisée à pH 7,0 où 90% du matériel protéique subsiste tout de même dans le surnageant de l'échantillon, au bout de 7 jours. Par contre, toute la gamme de pH située de 6,8 à 4 provoque la précipitation de la protéine. Même si une amélioration de la solubilité de la protéine est observable pour un pH de 3,8, les pH acides ne peuvent pas être utilisés lors de la première étape de purification de la protéine, sur colonne d'affinité au nickel. La possibilité d'utiliser ces pH acides pour la purification de SelA n'a donc pas été étudiée plus en détails.

Ces résultats montrent que les conditions dans lesquelles doit être conservée SelA sont très strictes, et que la meilleure des solutions est de conserver un tampon de même nature tout au long du protocole de purification, ce qui a été fait lors de la mise au point de celui-ci.

### 3.3.3.4 Evolution du protocole de purification

Je me suis donc intéressé plus en détail à l'art de la purification des protéines. Ma chance résidait en l'acquisition récente par le laboratoire d'un AKTA purifier<sup>®</sup>, système FPLC de précision, piloté par ordinateur (produit d'Amersham, aujourd'hui GE Healthcare). Comme les protocoles de purification de la plupart des protéines étudiées au laboratoire étaient déjà en place sur un système plus ancien (un GradiFrac<sup>®</sup> d'Amersham), il fut naturel que je m'investisse, pour la mise au point de la nouvelle purification de SelA, de la mission de la mise en route fonctionnelle du nouveau système.

Les premiers essais de purification de SelA sur colonne d'affinité avaient déjà été effectués sur cet appareil récent. La maîtrise et la compréhension de ses diverses fonctions m'apparaissait être un bon moyen de mieux contrôler les étapes de purification que j'allais devoir mettre au point sur cet appareil, et d'en comprendre les résultats.

Dans cette optique, j'ai pu assister à des formations sur l'utilisation avancée du système AKTA, mais également sur les stratégies de purification des protéines par chromatographie FPLC. La documentation existante sur le sujet, ainsi que les discussions avec des personnes d'autres laboratoires connaissant le système AKTA, m'ont permis de me forger une opinion sur la démarche à suivre.

La stratégie de purification qui m'a semblé la plus prometteuse est la combinaison de la première étape de purification de la protéine par affinité, avec une purification sur colonne de gel à exclusion stérique.

En effet, la taille importante de la forme décamerique de ecSelA et de mtSelA en solution (543 kDa) lui permet de se différencier facilement des autres protéines adsorbées non-spécifiquement sur la colonne d'affinité, qui contaminent les échantillons obtenus.

De plus, l'utilisation de cette technique de séparation peut être couplée à une calibration de la colonne : l'utilisation d'échantillons de référence de diverses tailles connues permet de relier le temps de rétention des composés à leur taille apparente.

Un troisième avantage de cette technique est la possibilité de changer la composition du tampon dans lequel se trouve la protéine sans avoir à passer par une étape de dialyse : quel que soit le tampon de l'échantillon injecté, la protéine sera éluée dans le tampon d'équilibrage de la colonne. Il est toutefois nécessaire de s'assurer au préalable de la stabilité de la protéine dans le tampon de purification.

J'ai donc mis en place sur l'AKTA purifier<sup>®</sup> du laboratoire une colonne Superdex200 et une colonne Superdex75, de 25mL de résine chacune, qui ont respectivement comme limite d'exclusion stérique 1300 kDa et 100 kDa, et comme plage de séparation résolutive de 10 à 600 kDa et de 3 à 70 kDa. La purification des protéines SelA en condition native s'effectue sur la colonne Superdex200.

La mise au point de programmes automatiques de nettoyage de la colonne, d'équilibrage de la colonne en tampon, d'injection de l'échantillon et de collecte des fractions d'intérêt, permet l'obtention de manipulations reproductibles et la comparaison objective des résultats obtenus lors de différentes purifications. Les programmes mis au point permettent également une utilisation aisée de ces colonnes, qui sont devenues des ressources communes utilisées aujourd'hui couramment par les membres du laboratoire intéressés par cette technique.

Les chromatogrammes obtenus lors de la purification ont permis d'accéder à beaucoup plus d'informations sur le comportement de SelA en solution que nous ne l'aurions présagé initialement.

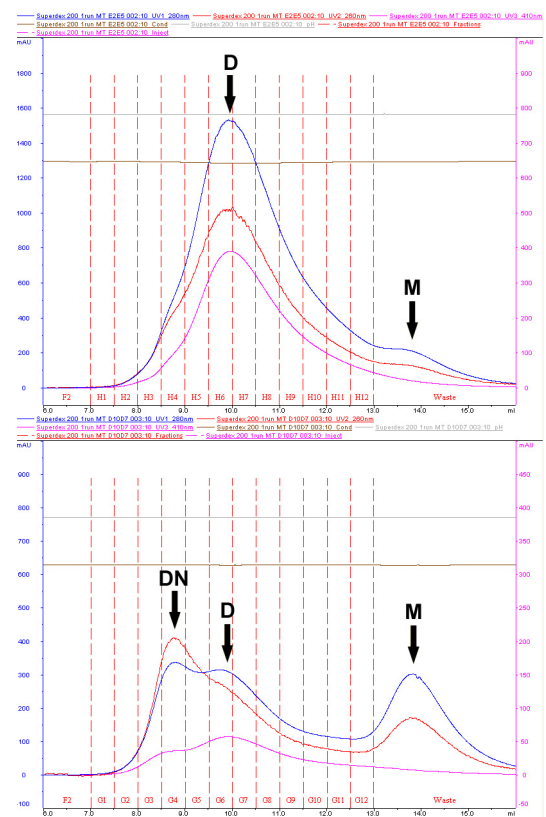
Lors de la purification sur la colonne de gel d'exclusion stérique Superdex200, le système AKTA permet le suivi de différents paramètres, en particulier l'absorption à 3 longueurs d'onde différentes, de l'échantillon à la sortie de la colonne. Le choix d'utiliser des longueurs d'onde de 260nm, 280nm et 410nm semble avoir été le plus approprié. Les deux premières longueurs d'onde sont en effet les choix classiques de détection de la présence de matériel nucléique et protéique, respectivement. La longueur d'onde de 410nm est, elle, caractéristique du cofacteur de l'enzyme, le pyridoxal-5'-phosphate, qui donne une couleur jaune aux solutions de SelA.

Le suivi spectrométrique des purifications permet l'observation des points suivants :

- La protéine SelA est éluée hors de la colonne à un volume de 9,9 mL après l'injection de l'échantillon (voir chromatogramme du haut de la figure 3.9). Ce volume de rétention correspond à la masse attendue pour le décimère d'environ 540 kDa (la Ferritine de 440 kDa est éluée à 10,9 mL et la Thyroglobuline de 669 kDa est éluée à 9,4 mL, l'échelle de masse n'étant bien sûr pas une fonction linéaire du volume d'élution).

La présence en solution d'ecSelA et de mt-SelA sous forme de décimère est une confirmation du fait que le Tag à 6 histidines ne gêne pas la formation de la structure oligomérique classique de la protéine. Mieux, la protéine tronquée mtSelA( $\Delta$ 1-31) forme également un décimère en solution, ce qui montre que l'extrémité N-terminale n'est pas utile à l'organisation oligomérique dans le cas de SelA, contrairement à SepSecS (voir § 2.3.5.2, page 100).

- Bien que la majeure partie de la protéine SelA appartienne à une population homogène de décimère (notée D), une proportion variable de SelA recombinante sur-exprimée dans *E. coli* se trouve sous des formes aux caractéristiques différentes. Une population de monomères (notée M), d'importance variable, peut être observée lors de la purification sur la colonne de filtration sur gel (voir le chromatogramme du bas de la figure 3.9), avec un volume de rétention de 13,9 mL correspondant à la masse apparente de 54 kDa (l'Ovalbumine de 43 kDa est éluée à 14,7 mL et l'Albumine de 67 kDa est éluée à 13,4 mL). Des



**Figure 3.9** – Chromatogrammes de purification de mtSelA sur la colonne Superdex200. (260nm en rouge, 280nm en bleu, 410nm en rose)

échantillons des populations de Sela monomériques et décamériques sont indiscernables sur gel SDS PAGE. La population M est constituée de protéines ne possédant pas de cofacteur PLP, vu leur spectre d'absorption à 410nm. Il s'agit sans aucun doute de protéines dont le repliement n'est pas standard, ne permettant ni la formation de décamères, ni la conservation du cofacteur PLP, dont le tampon de lyse est pourtant pourvu, à hauteur de 4 $\mu$ M en accord avec les publications de référence.<sup>[II.9, II.10, II.14]</sup>

Cette population se retrouve en proportion plus importante dans les premières fractions de la purification sur colonne d'affinité. Il peut donc en être conclu que son état ne favorise pas une forte affinité avec les ions nickel de la colonne.

- Une troisième population de protéines Sela (notée DN) est éluée à des volumes variables, entre 8,4 mL et 8,8 mL. Cette population de protéines présente une absorption à 260nm supérieure à la normale, voire plus importante que son absorption à 280nm (voir le chromatogramme du bas de la figure 3.9, page précédente). Cette population est rencontrée également dans une plus grande proportion dans les premières fractions de purification sur colonne d'affinité.

Vu la masse apparente qui caractérise cette population, et l'importance de son absorption à 260nm — qui n'est cependant pas au niveau de celle du complexe de Sela et de l'ARN<sub>t</sub><sup>Sec</sup> (voir § 3.4, page 153, dédié à la formation du complexe) — la population DN est sans doute composée de décamères de Sela interagissant de manière non-spécifique avec du matériel nucléique, qui provient de la lyse cellulaire, et formant des agrégats, composés d'au moins deux décamères. Là aussi, il n'y a pas de différence visible entre des échantillons des populations D et DN sur un gel SDS PAGE coloré au bleu de Coomassie, selon le procédé habituel.

Des agrégats au volume d'éluion similaire sont également observables lors du vieillissement d'une solution de protéine Sela, ou dans le surnageant d'une solution ayant entamé la phase de précipitation. Dans ce cas, cependant, l'absorption à 260nm reste inférieure à celle à 280nm, dans les normes caractérisant le matériel protéique.

L'existence de la population DN permet de tirer une conclusion importante : la protéine Sela possède bien des sites de surfaces affins pour du matériel nucléique chargé négativement. Par contre, les interactions prenant place au niveau de ces sites paraissent non-spécifiques.

- Les statistiques que permet d'effectuer l'accumulation des données spectrométriques sur ces différentes populations de protéines lors des différentes sessions de purification, donnent l'accès à des observations intéressantes. Par exemple, il a été noté que le rapport entre l'absorption à 410nm et à 280nm était variable entre les différentes sessions de purification effectuées, pour la population de décamère d'intérêt.

Cela montre que la quantité de PLP présente dans un échantillon n'est pas proportionnelle à la quantité d'enzyme Sela de l'échantillon, dans nos conditions de purification. J'ai donc décidé d'effectuer des essais de complémentation de Sela en PLP, pour vérifier le statut de l'enzyme vis-à-vis de son cofacteur, essentiel pour la formation du complexe avec le seryl-ARN<sub>t</sub><sup>Sec</sup>.

### 3.3.3.5 L'ajout de cofacteur

Les protocoles de purification des Sela<sub>sauvage</sub> proposés dans la littérature<sup>[II.9,II.10,II.14]</sup> suggèrent l'emploi d'un tampon contenant un peu de PLP (4 à 10 $\mu$ M) lors d'une des étapes de purification de Sela.

Cette méthode n'est pas suffisante dans notre cas, en plus d'être consommatrice en PLP. Pour le montrer, et tester une méthode alternative de complémentation de Sela en cofacteur, un échantillon de purification sur colonne d'affinité est séparé en deux volumes égaux. Ces derniers sont laissés une nuit entière à 4°C, l'un sans traitement, l'autre avec un ajout de PLP jusqu'à obtenir une concentration finale de 100 $\mu$ M.

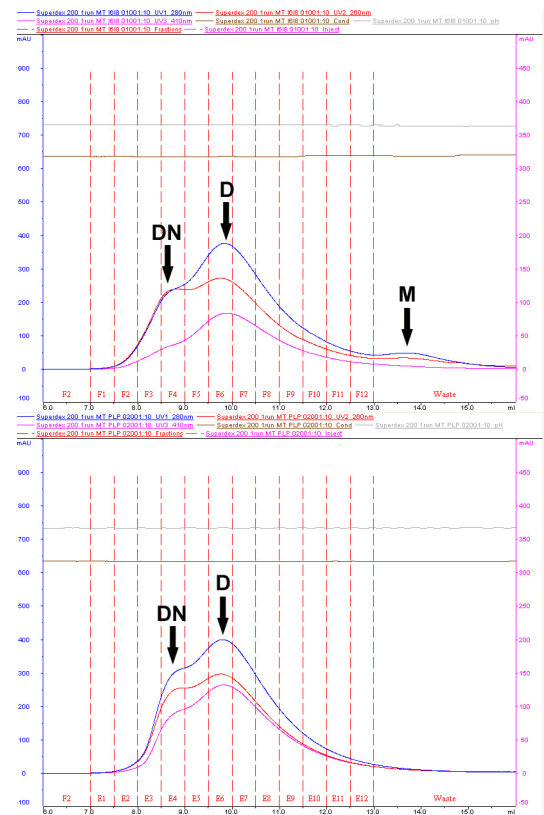
Une purification des deux échantillons sur la colonne de gel à exclusion stérique habituelle permet d'obtenir les chromatogrammes présentés en figure 3.10, l'échantillon non traité apparaissant en haut, et l'échantillon traité en bas.

L'augmentation de la proportion de PLP par protéine Sela est nettement visible, et est étonnamment plus importante pour la population DN que pour la population d'intérêt D (voir l'augmentation de l'absorption à 410nm (en rose) sur les chromatogramme de la figure 3.10). Il est également observable que la population M de protéine monomérique est inexistante dans l'échantillon traité. Cet échantillon présentait un léger précipité. Il est impossible de savoir si ce précipité était constitué de la population M ou si la complémentation en PLP de l'échantillon a permis à ces monomères de complexer une molécule de cofacteur, de se restructurer autour de ce centre catalytique et de former des décimères qui ont rejoint les populations D et DN.

La stratégie de purification finalement utilisée prend en compte cette nécessité de complémenter l'échantillon en PLP.

### 3.3.3.6 Conclusions sur la purification de Sela

Le protocole de purification mis au point est donc constitué de 3 étapes : une purification du lysat cellulaire sur colonne d'affinité, une complémentation des échantillons en PLP durant une nuit, puis une seconde purification sur colonne de filtration sur gel Superdex200. Cette dernière permet d'éliminer les impuretés résiduelles et l'excédant de PLP ainsi que de changer le tampon de Sela pour celui dans lequel sera effectuée la formation du complexe avec l'ARN<sub>t</sub><sup>Sec</sup>.



**Figure 3.10** – Effet d'une complémentation massive en cofacteur PLP.



En conclusion, la technique de purification sur gel à exclusion stérique permet d'obtenir, en plus de données essentielles sur leur comportement en solution, les protéines ecSelA, mtSelA et mtSelA( $\Delta$ 1-31) dans des conditions contrôlées, en des temps relativement courts et à un très bon niveau de pureté. Aucun échantillon obtenu par cette technique de purification ne présente d'activité RNase.

La figure 3.11 illustre bien le gain de pureté obtenu grâce à cette technique : les deux puits de gauche contiennent des échantillons issus d'une purification sur colonne d'affinité, concentrés pour souligner les impuretés conservées lors de cette première étape de purification. Les trois puits de droite contiennent des fractions de purification sur la colonne Superdex200 de ces mêmes échantillons. Il subsiste une contamination résiduelle visible, de poids supérieur au monomère de SelA, mais de nature inconnue. Les traces de composés de faibles poids moléculaires, colorés de couleur différente, ont été identifiés comme étant des oligo de petites tailles, interagissant non-spécifiquement avec SelA, comme ceux à l'origine de la formation de la population DN de haut poids moléculaire.



**Figure 3.11** – Gel SDS PAGE coloré à l'argent, d'échantillons purifiés de mtSelA.

Les échantillons obtenus par cette méthode sont suffisamment purs pour pouvoir être utilisés dans les expériences de mesures biochimiques et dans les essais de formation du complexe et de cristallogénèse. Pour ces manipulations sensibles, j'ai préféré utiliser à chaque nouvel essai des échantillons fraîchement purifiés, et donc plus homogènes.

Pour pouvoir effectuer ces manipulations, il est également nécessaire de connaître avec précision la concentration de la protéine dans les échantillons obtenus, et donc de disposer des coefficients d'absorption réels des protéines utilisées. Dans notre cas, la présence du cofacteur PLP ne nous permet pas d'utiliser les coefficients d'absorption théoriques de la protéine, le pyridoxal-5'-phosphate ayant, en plus de son absorption à 410nm, des coefficients d'absorption importants à 280nm et à 260nm.

Protéine	$\epsilon_{(th)}$ à 280nm	$\epsilon_{(exp)}$ à 280nm	$\epsilon_{(exp)}$ à 260nm	$\epsilon_{(exp)}$ à 410nm
ecSelA	35410 M <sup>-1</sup> .cm <sup>-1</sup>	39500 M <sup>-1</sup> .cm <sup>-1</sup>	29200 M <sup>-1</sup> .cm <sup>-1</sup>	8700 M <sup>-1</sup> .cm <sup>-1</sup>
mtSelA	25900 M <sup>-1</sup> .cm <sup>-1</sup>	29300 M <sup>-1</sup> .cm <sup>-1</sup>	22500 M <sup>-1</sup> .cm <sup>-1</sup>	8500 M <sup>-1</sup> .cm <sup>-1</sup>

**Tableau 3.3** – Coefficients d'absorption de SelA. Les coefficients  $\epsilon_{(exp)}$  ont été mesurés expérimentalement à pH 7,8.

La collecte des données spectroscopiques de purification de SelA a permis de vérifier les calculs des absorptions molaires obtenues par dosage de solutions données de protéines

ecSelA et mtSelA par la méthode de Bradford. Ces dernières sont présentées dans le tableau 3.3, page ci-contre.

## 3.4 Formation du complexe

Les différentes versions de SelA et du séryl-ARN<sub>t</sub><sup>Sec</sup> ainsi obtenues nous permettent de former différents complexes et de tenter de les cristalliser.

Pour des raisons chronologiques et de plus grand rendement de transcription, le séryl-ARN<sub>t</sub><sup>Sec</sup> d'*Escherichia coli* est l'ARN<sub>t</sub> qui a été utilisé dans la plupart des expériences de complexation, et cela avec ecSelA, mtSelA et mtSelA( $\Delta$ 1-31). L'ARN<sub>t</sub><sup>Sec</sup> de *Moorella thermoacetica* a en effet été produit plus difficilement, en petites quantités, et a donc surtout été réservé pour des études d'interactions.

### 3.4.1 Mise en présence des partenaires

La stratégie que nous avons utilisé pour former le complexe entre le séryl-ARN<sub>t</sub><sup>Sec</sup> et ecSelA dans nos premiers essais de cristallisation, est un ajout direct de la solution concentrée de séryl-ARN<sub>t</sub><sup>Sec</sup> dans une solution de tampon adéquat à pH 7,0 contenant l'enzyme SelA diluée, les deux partenaires étant utilisés dans leur proportion stœchiométrique (1 mole d'ARN<sub>t</sub> pour 2 moles de SelA). La solution obtenue est alors concentrée au taux désiré (entre 5 et 10 mg/mL) pour être utilisée en cristallogénèse.

Là aussi, la stratégie employée a évolué, grâce à l'accès à la purification sur gel à exclusion stérique et à une meilleure compréhension de la formation du complexe.

Les conditions de formations du complexe génèrent, en effet, une problématique à laquelle il a été difficile de trouver la solution la plus adéquate, car :

- Le séryl-ARN<sub>t</sub><sup>Sec</sup> est une espèce se déacylant facilement. Il doit être conservée à pH acide (5,0 à 5,5), comme le montrent les résultats de cinétique de déacylation, exposés au § 4.2, page 179.
- Les mesures de stabilité de SelA, exposées au § 3.3.3.3, page 147, montrent que la protéine est plus stable à pH basique qu'à pH neutre et acide. Une brusque descente en pH, qui peut être provoquée par l'ajout de l'ARN<sub>t</sub><sup>Sec</sup> et de son tampon acide, provoque l'agrégation et la précipitation de la protéine. Les deux entités de formation du complexe peuvent donc difficilement être conservées dans un tampon identique, compatible avec un mélange homogène de deux solutions de même tampon lors de la réaction de complexation, méthode de prédilection pour ce genre de réaction.
- Par contre, une fois formé, le complexe perd le caractère instable de SelA, prompt à l'agrégation, même avec un faible chargement des décamères en ARN<sub>t</sub><sup>Sec</sup>. De même, le séryl-ARN<sub>t</sub><sup>Sec</sup> se déacyle moins rapidement lorsqu'il est en complexe avec SelA (voir § 4.2, page 179).

La solution à ce problème, trouvée grâce aux essais de complexation suivis par purification "analytique" (filtration sur gel), réside en l'ajout successif des deux solutions — SelA dans son tampon HEPES à pH 7,8 et le séryl-ARN<sub>t</sub><sup>Sec</sup> dans son tampon acétate à

pH 5,5 — à une solution de réaction composée de tampon HEPES à pH 7,0 de concentration élevé (jusqu'à 500mM), et donc à haut pouvoir tampon. La fluctuation du pH dans la solution de réaction est alors inexistante, ce qui garantit une certaine stabilité transitoire aux deux entités, réagissant pour constituer le complexe, forme sous laquelle elles sont toutes les deux immobilisées.

Un autre problème a été identifié grâce au suivi chromatographique de la formation du complexe : si la concentration de Sela dépasse 15 $\mu$ M dans le milieu réactionnel, de la formation du complexe est moindre. Cela est sans doute lié à la tendance à l'agrégation de Sela, mais il est difficile de la vérifier.

Le résultat d'une réaction suivant ces recommandations est l'obtention importante de complexe, comme l'illustre le chromatogramme du bas de la figure 3.12. Dans le cas de l'ajout du tampon acide contenant le séryl-ARN<sub>t</sub><sup>Sec</sup> dans une solution de Sela, pourtant tamponnée à pH 7,8, produit une agrégation de la protéine, alors incapable de se lier à l'ARN<sub>t</sub> introduit dans le milieu (chromatogramme du haut de la figure 3.12).

Les trois pics massifs visibles sur ces chromatogrammes sont :

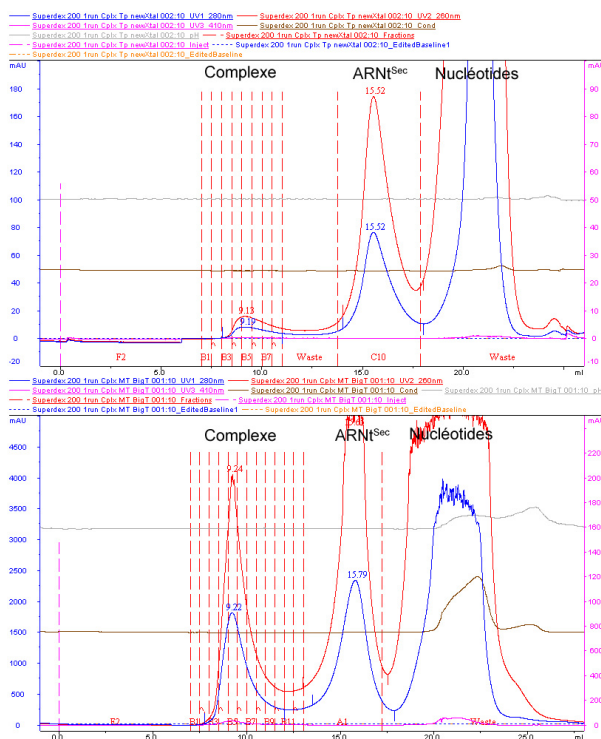
- le complexe Sela/séryl-ARN<sub>t</sub><sup>Sec</sup> élué entre 9,0 et 9,3 mL ;
- le séryl-ARN<sub>t</sub><sup>Sec</sup> et l'ARN<sub>t</sub><sup>Sec</sup> déacétylé, élués entre 15,5 et 16,2 mL ;
- les nucléotides ATP et AMP, provenant de la réaction de charge en sérine de l'ARN<sub>t</sub><sup>Sec</sup> et co-précipités avec lui lors de sa purification, élués entre 20 et 24 mL.

Deux éléments importants découlent des informations présentées sur ces chromatogrammes :

- Le premier est la confirmation de la formation du complexe stable désiré entre l'ARN<sub>t</sub><sup>Sec</sup> d'*E. coli* et les protéines ecSela et mtSela. La formation du complexe avec la protéine tronquée mtSela( $\Delta$ 1-31) n'a pas été testée dans ces conditions.
- Le deuxième est une confirmation de la possibilité de purifier le complexe formé entre nos deux partenaires : la colonne de gel à exclusion stérique peut être utilisée à cette fin, ce qui est le cas dans l'exemple de la figure 3.12, chromatogramme du bas.

### 3.4.2 Purification du complexe

Cette purification permet d'obtenir un échantillon de complexe dans un tampon de cristallisation de composition choisie, avec lequel la colonne doit simplement être équilibrée au préalable. Ainsi, le complexe est débarrassé des sels indésirables ou en quantité trop



**Figure 3.12** – Étude sur la formation du complexe entre le séryl-ARN<sub>t</sub><sup>Sec</sup> d'*E. coli* et mtSela.

importante, des composés contaminant comme les nucléotides importés avec le séryl-ARN<sub>t</sub><sup>Sec</sup>, de l'excès de séryl-ARN<sub>t</sub><sup>Sec</sup> et d'ARN<sub>t</sub><sup>Sec</sup> déacylé, qui peuvent tous deux être récupérés pour être recyclés.

Grâce à cette technique, la quantité de séryl-ARN<sub>t</sub><sup>Sec</sup> apportée lors de la réaction peut être augmentée jusqu'à des proportions sensées permettre d'obtenir un complexe entièrement chargé — 5 séryl-ARN<sub>t</sub><sup>Sec</sup> par décamère — c'est à dire jusqu'à un ratio de séryl-ARN<sub>t</sub><sup>Sec</sup>/SelA de 3 pour 1, soit 6 fois le rapport stœchiométrique.

La charge réelle du complexe obtenu n'est pas mesurable : la seule méthode disponible est spectroscopique, et le changement de pH de 7,8 à 7,0, le changement de liaison du PLP et surtout les interactions entre l'ARN<sub>t</sub><sup>Sec</sup> et SelA, font des coefficients d'absorption du complexe des inconnues indéterminables. Les calculs effectués sur la base des coefficients d'absorption connus des différents partenaires à l'état libre et sur le rapport observable entre l'absorption à 280nm et à 260nm du complexe donnent un ratio d'environ 4 ARN<sub>t</sub><sup>Sec</sup> pour un décamère, ratio sans valeur formelle.

Ainsi, la population de complexe créée et purifiée, le plus homogène possible, est prête pour une concentration et une utilisation en cristallogénèse.

### 3.4.3 Essais de production directe du complexe

Une observation lors de la purification de SelA a particulièrement éveillé notre intérêt : la population DN observable en purification est le produit de l'interaction non-spécifique entre les décamère SelA et du matériel nucléique de taille modérée provenant de la lyse cellulaire. Cette interaction est suffisamment forte pour que les complexes non-spécifiques formés ne se dissocient pas durant la filtration sur gel.

Si la souche d'*Escherichia coli* BL21 DE3 employée sur-exprimait l'ARN<sub>t</sub><sup>Sec</sup> en plus de la protéine SelA, obtiendrait-on un complexe suffisamment stable entre ces deux partenaires spécifiques pour pouvoir les purifier, et laisser ainsi le soin à l'usine bactérienne de produire pour nous nos macromolécules d'intérêt ?

La technique de production d'ARN<sub>t</sub> en bactéries recombinantes est connue,<sup>[II.15-II.19]</sup> mais l'intérêt de la double production des partenaires protéique et nucléique que nous envisagions était de simplifier la purification de l'ARN<sub>t</sub><sup>Sec</sup> produit, ce qui est généralement l'étape laborieuse du procédé.

Nous avons donc transformé une souche d'*Escherichia coli* BL21 DE3 avec le plasmide pET29b contenant le système de sur-expression de ecSelA, et avec le plasmide pSTV28 contenant le système de sur-expression modérée de l'ARN<sub>t</sub><sup>Sec</sup> d'*E. coli*.

Une sélection des clones contenant les deux plasmides est effectuée grâce à leur double résistance aux antibiotiques chloramphenicol et kanamycine.

À la suite d'une culture classique de clones sélectionnés, le lysat cellulaire est soumis à la même stratégie de purification que la protéine seule. Lors de la première étape de purification sur le système FPLC — sur colonne d'affinité au nickel — la protéine est

éluée comme à l'habitude, avec une signature spectrale classique de matériel protéique, sans matériel nucléique. Une vérification par une purification sur colonne de filtration sur gel confirme l'absence de matériel nucléique complexé à SelA, le profil spectroscopique de l'élué ne différant pas d'une purification classique de la protéine.

Le problème de cette méthode est sans doute l'utilisation du système FPLC pour la première étape de purification sur résine d'affinité au nickel : le lavage en continu, par du tampon d'élué, de la résine sur laquelle sont fixés les décamères de SelA, a dû favoriser la perte du matériel nucléique que nous souhaitions récupérer, c-à-d l' $ARN_t^{Sec}$ .

J'ai donc effectué d'autres essais de purification du lysat de cellules contenant les deux plasmides, en particulier par lots (ou batch) dans des tubes contenant de la résine d'affinité au nickel. Les lavages manuels de la résine avec les tampons de purification, suivis d'un temps de repos puis d'une centrifugation, devaient permettre aux partenaires de rester fixés à la résine. Malheureusement, les vérifications de la contenance des échantillons ainsi purifiés sur gel de polyacrylamide montrent que la technique de purification par lots n'est pas suffisamment efficace : une population très importante de matériel nucléique est retenue, sans qu'il y ait d'enrichissement notable de ce matériel en  $ARN_t^{Sec}$ .

Étant donné que les purifications des deux partenaires  $ARN_t^{Sec}$  et SelA, déjà effectuées séparément, permettaient de former facilement le complexe désiré, cette tentative d'innovation par une méthode élégante de production et de purification d'un ARN grâce à son interaction avec son partenaire protéique co-exprimé fut abandonnée.

### 3.5 Cristallogenèse

Une fois les deux partenaires produits et purifiés, leur complexe formé et purifié, l'étape principale nécessaire à l'obtention de données structurales sur notre objet d'étude peut être entamée.

La cristallogenèse est une technique largement employée en biologie structurale. Beaucoup de laboratoires se sont spécialisés dans la maîtrise de cet outil, et l'utilisent comme technique pivot appliquée à divers objets d'études : protéines cytoplasmiques ou membranaires, acides nucléiques, complexes protéiques, complexes protéo-nucléiques et virus.

Dans le cas de notre laboratoire, l'utilisation de cette technique est récente, et fut appliquée dans deux projets en plus du présent travail. Ces projets ont débouché sur la résolution de 3 structures de complexes protéine-ARN (code d'accès PDB 1WSU, 2PJP, 2PLY). Les sujets d'études étaient, dans un cas, les complexes entre le domaine C-terminal de SelB et un élément SECIS d'un  $ARN_m$  pour les organismes *Escherichia coli*<sup>[II.20,II.21]</sup> et *Moorella thermoacetica*<sup>[II.21,II.22]</sup> (complexes 1 pour 1), travail de doctorat de Nicolas Soler,<sup>[II.23-II.25]</sup> et dans l'autre cas, le complexe entre la méthyltransférase RlmAII et le fragment d'ARN ribosomal qu'elle modifie chez *Streptococcus pneumoniae* (complexe 1 pour 1), travail de doctorat de Clotilde Husson.<sup>[III.26,II.27]</sup>

Pour mon étude, j'ai utilisé les mêmes méthodes de cristallisation que mes collègues, en appuyant mon apprentissage des techniques utilisées sur leurs connaissances et sur les discussions tenues avec le personnel d'autres laboratoires employant également des techniques similaires.

### 3.5.1 Techniques employées

La cristallisation de macromolécules est une technique empirique, pour laquelle le comportement observé diffère beaucoup d'un cas à l'autre. Malgré, cet indéterminisme qui ne permet pas de connaître à l'avance les conditions dans lesquelles une macromolécule formera des cristaux, la cristallisation suit certains principes théoriques généraux.

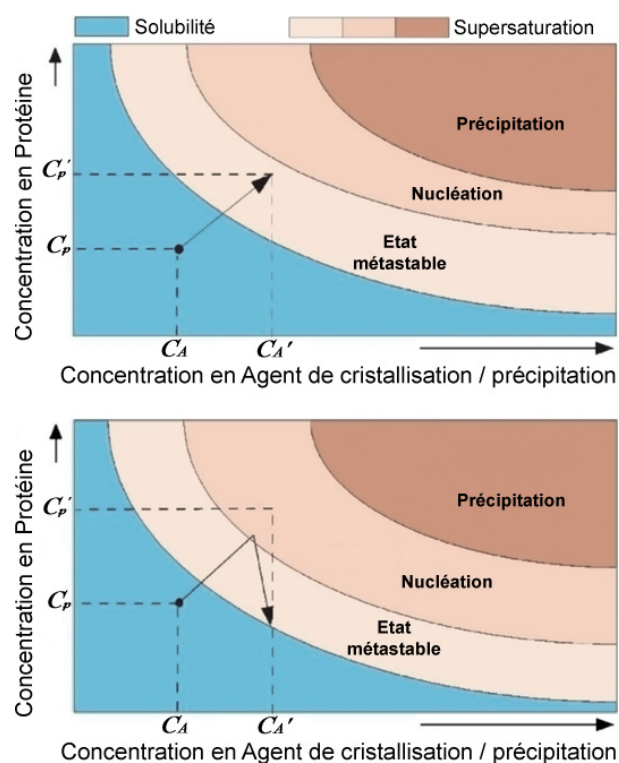
#### Principes de la cristallisation de macromolécules

La technique de cristallisation dont l'utilisation est la plus répandue dans la recherche structurale, est la cristallisation par diffusion de vapeur.

Que ce soit par goutte suspendue ou par goutte assise (voir paragraphes suivants), cette technique est basée sur la concentration d'une solution, mélange de l'échantillon protéique et d'agents de précipitation. Cette concentration est induite par l'échange de vapeur qui s'opère entre la goutte du mélange et un réservoir contenant les agents de précipitation à une concentration deux fois plus élevée, jusqu'à équilibrage des tensions de vapeur des deux solutions.

Dans la goutte de mélange, la perte d'eau par échange de vapeur permet de faire passer la protéine d'un état soluble à un état de sursaturation, qui, selon le diagramme des phases de cette macromolécule dans la solution de précipitant donnée, peut faire partie d'un domaine d'état métastable, de nucléation ou de précipitation de la protéine.

Tout le travail du cristallographe est de trouver les conditions initiales (la concentration de la macromolécule, la composition du mélange de précipitants, le pH, la température, la force ionique, l'effet d'additifs) permettant une évolution vers



**Figure 3.13** – Diagramme des phases d'une protéine et variation de l'état dans le cas de la diffusion de vapeur. **(En haut)** La protéine reste soluble, dans un état métastable de supersaturation, au-delà de la limite de solubilité. **(En bas)** Lorsque la protéine atteint un état de nucléation, des cristaux peuvent se former et appauvrir la solution en protéines lors de leur croissance.

l'état de nucléation (voir figure 3.13). Cette recherche dans l'espace multidimensionnel des variables de cristallisation est souvent l'étape la plus consommatrice de temps dans le processus menant à l'obtention de cristaux.

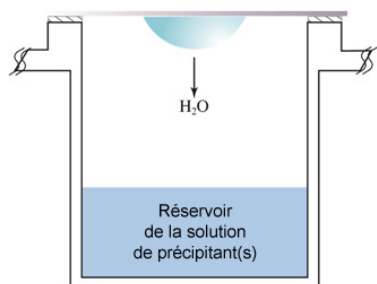
Pour faciliter cette tâche, une multitude de kits de conditions de cristallisation sont commercialement disponibles : leurs compositions explorent l'ensemble des possibilités, de façon clairsemée (compilations de solutions connues, plans factoriels clairsemés à multiples paramètres) ou dans certaines directions (par exemple, kit balayant l'espace des pH et des additifs pour un certain précipitant).

Parmi les conditions qui amènent à un état de nucléation de la protéine, seules celles possédant le bon compromis entre facteurs thermodynamiques (disposition du diagramme de phase, solubilité et précipitation) et facteurs cinétiques (potentiel de nucléation et vitesse de croissance cristalline) aboutissent à une cristallisation de la macromolécule sous une forme intéressante : si possible, de grands cristaux homogènes et solitaires, qui permettront d'effectuer des mesures de diffractions aux rayons X.

En général, de légères modifications de la condition de cristallisation d'intérêt permettent de s'approcher de ce but.

### Méthode de la goutte suspendue

Dans le cadre de la technique de diffusion de vapeur, la méthode de la goutte suspendue est la méthode manuelle de cristallogénèse la plus populaire.



**Figure 3.14** – Puits de cristallogénèse utilisant la méthode de la goutte suspendue.

Le principe en est le suivant :

- La plate-forme de travail est constituée, dans le cadre de notre utilisation, d'une plaque de 24 puits.
- Le réservoir de chaque puits est prévu pour accueillir un volume recommandé de 500  $\mu$ L à 1mL de la condition de cristallisation, c-à-d. de la solution de précipitants.
- Une lamelle ronde en verre silicé est le support sur lequel la goutte de cristallisation sera positionnée. Cette dernière est obtenue en mélangeant un volume égal (selon les cas) de solution contenant la macromolécule à cristalliser et de la condition de cristallisation présente dans le réservoir. Le volume de cette goutte peut varier

- de 1  $\mu$ L à 40  $\mu$ L. Des additifs peuvent également être ajoutés à la goutte de cristallisation.
- La lamelle de verre est retournée, positionnée au-dessus du puits, où un anneau de graisse à vide de silicone permet de sceller le puits et de laisser libre cours à la diffusion de vapeur entre la goutte et le réservoir.

Cette méthode allie les avantages suivants :

- L'observation des gouttes de cristallisation est facilitée par leur présence sur la lamelle de verre scellant le puits.
- L'accès aux cristaux est aisé : l'ouverture du puits permet de disposer de la lamelle portant la goutte, pour une "pêche" aux cristaux facilitée.

- Elle permet l'utilisation de gouttes de volume important, ce qui aboutit théoriquement à l'obtention de plus gros cristaux lorsque les conditions de cristallisation sont cernées.
- Elle permet de choisir manuellement les conditions utilisées, d'effectuer des mélanges entre solutions, d'utiliser des solutions de précipitants personnalisées, d'ajouter des additifs à la goutte de cristallisation.
- Plusieurs gouttes peuvent être placées sur la même lamelle de verre, pour une même condition de cristallisation. Cela est souvent utile lorsque plusieurs concentrations de macromolécules sont testées, ou pour observer l'effet du volume de la goutte sur la forme des cristaux obtenus, ou encore pour comparer les effets de différents additifs.

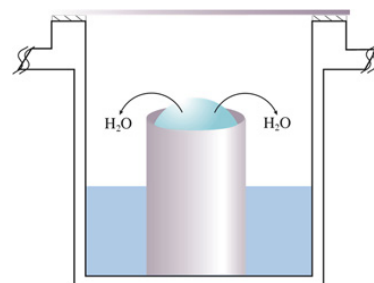
En raison de sa facilité de mise en place, cette méthode est généralement celle des premières expériences de cristallogenèse pour un projet donné. Elle permet ainsi d'accéder aux premières observations sur le comportement de la macromolécule face à certains précipitants, à certains sels ou à certains pH.

C'est donc avec cette méthode que j'ai effectué les premiers essais de cristallisation de SelA et du complexe SelA / séryl-ARN<sub>t</sub><sup>Sec</sup>. C'est également la méthode que j'ai utilisée lors des essais de reproductibilité de conditions de cristallisation avérées, dans le but d'obtenir des cristaux utilisables pour la détermination de structures par diffraction aux rayons X. (voir § 3.5.2, page suivante, pour un détail des expériences de cristallogenèse effectuées.)

### Méthode de la goutte assise

Le plus grand intérêt de la méthode de la goutte assise est qu'elle peut être automatisée : des robots pipeteurs sont conçus pour prélever successivement la solution de macromolécules et la solution de cristallisation du réservoir, et les déposer à l'emplacement destiné à accueillir la goutte assise, où elles se mélangent.

J'ai eu la chance de pouvoir accéder à de tels robots, grâce à l'ouverture de la plate-forme de cristallisation automatisée du LEBS (Laboratoire d'Enzymologie et Biochimie Structurales, Gif-sur-Yvette) aux autres laboratoires du campus, à partir de l'été 2006.



**Figure 3.15** – Puits de cristallogenèse utilisant la méthode de la goutte assise.

Les caractéristiques de cette méthode sont les suivantes :

- Les plaques utilisées sont des micro-plaques de 96 puits, ayant chacun trois emplacements pour un dépôt de goutte.
- Les réservoirs sont remplis de 100  $\mu$ L de la solution de précipitant.
- Les gouttes peuvent avoir un volume de 0,1  $\mu$ L à 0,3  $\mu$ L, avec la possibilité d'ajouter automatiquement une solution d'additifs.
- Après la répartition, par le robot, des gouttes de macromolécules puis de conditions de cristallisation dans leurs emplacements pour chaque puits, la plaque est scellée par l'opérateur à l'aide d'un large ruban adhésif transparent, spécialement prévu à cet effet.



Cette méthode a pour avantage :

- la rapidité de l'exécution robotisée, qui permet l'obtention d'une plaque de 96 conditions de cristallisation en 45 minutes (les réservoirs doivent être remplis au préalable) ;
- le faible volume d'échantillon de macromolécule utilisé par puits, qui permet du coup de tester un nombre beaucoup plus important de conditions de cristallisation avec un volume d'échantillon donné,
- l'utilisation d'un scellage des puits sans graisse à vide, qui peut se révéler problématique dans certains cas.

Bien sûr, cette méthode possède également des inconvénients importants faisant qu'elle ne convient pas à tous les essais de cristallogénèse. En particulier, les faibles volumes utilisés permettent d'étudier les conditions dans lesquelles la macromolécule forme des cristaux, mais ceux-ci sont rarement de taille intéressante, et dans le cas contraire, leur extraction du dispositif à goutte assise est ardue.

Cette méthode est donc utilisée pour effectuer un premier criblage assez large des conditions de cristallisation, et réaliser des premiers essais d'optimisation des conditions, dans lesquelles un état de nucléation a pu être obtenu.

### 3.5.2 Carnet de bord de la cristallogénèse

Voici le détail chronologique des essais de cristallogénèse effectués durant mon doctorat. Cette vision globale ordonnée permet de situer les différentes étapes de ce travail, jalonné d'améliorations de protocoles et d'accès à de nouvelles techniques, qui ont été abordées tout au long de ce chapitre.

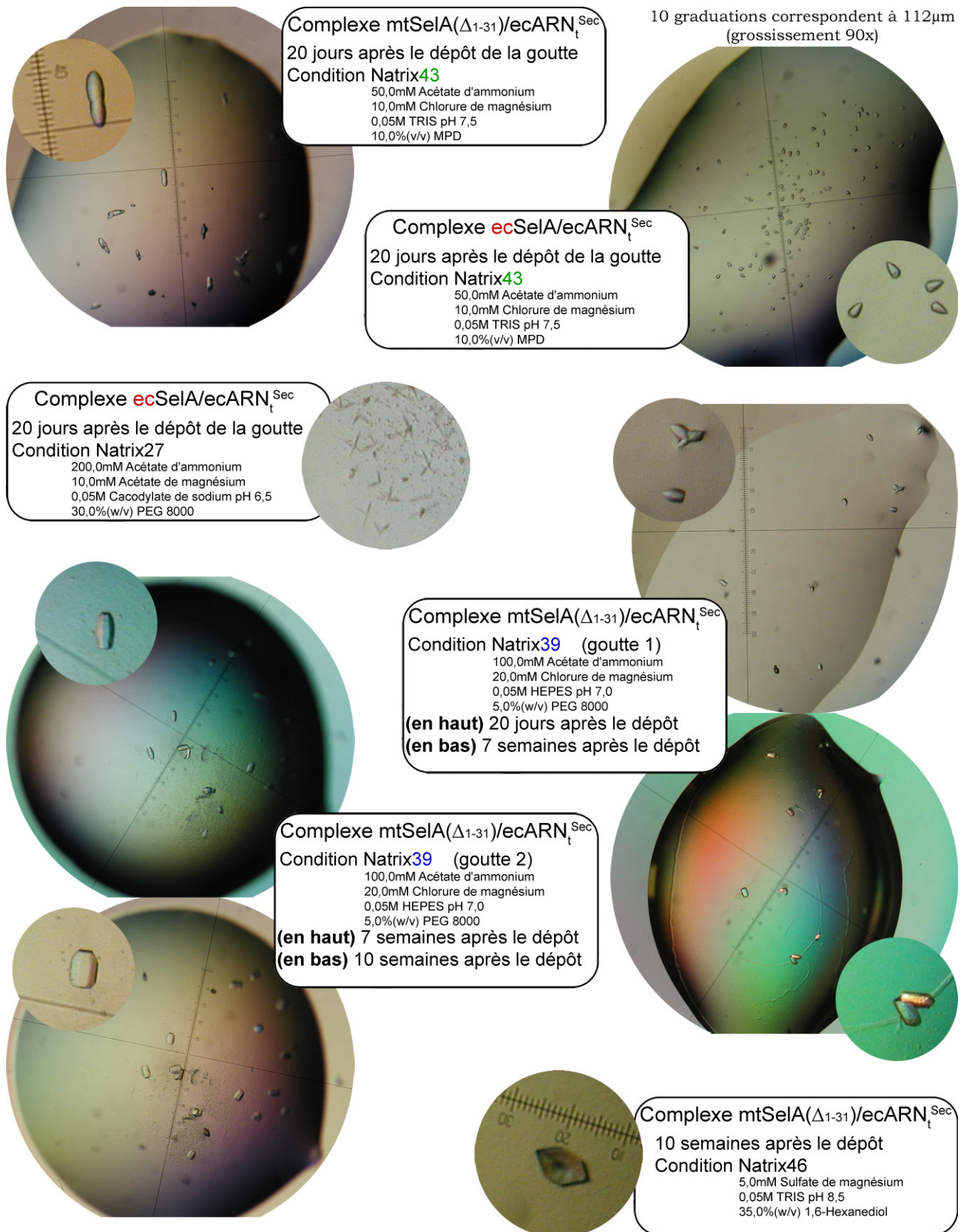
#### Première approche de la technique de cristallisation

Au bout de deux mois de travail sur le sujet, les premiers lots de mtSelA( $\Delta$ 1-31), de ecSelA et du séryl-ARN<sub>t</sub><sup>Sec</sup> d'*E. coli* voient le jour, en petite quantité, mais suffisante pour qu'ils soient utilisés dans un premier essai de formation et de cristallisation des deux complexes possibles.

(28/10/2005)	Cristallogénèse n° 1
98 conditions	kits Natrix et Cristal Screen I de Hampton Research®
Gouttes suspendues, mise en place manuelle	
Espèces mises en oeuvre	Complexe ecSelA / ARN <sub>t</sub> <sup>Sec</sup> d' <i>E. coli</i> Complexe mtSelA( $\Delta$ 1-31) / ARN <sub>t</sub> <sup>Sec</sup> d' <i>E. coli</i>
Formation du complexe	mélange simple ARN <sub>t</sub> <sup>Sec</sup> / SelA (1:2)

**Tableau 3.4** – Résumé de la cristallogénèse n° 1.

Lors de ce premier essai de cristallogénèse, des cristaux apparaissent dans certaines conditions de cristallisation au bout d'une semaine de diffusion de vapeur. (voir les photos de la figure 3.16, page suivante)



**Figure 3.16** – Morceaux choisis de la Cristallogenèse n°1. On peut voir, en haut, la différence entre les cristaux obtenus dans une même condition (Natrix43) pour le complexe à base d'ecSela et de mtSela( $\Delta_{1-31}$ ), différentes morphologies de cristaux selon les conditions (Natrix27, 39, 43 et 46), et la croissance des cristaux dans le temps, selon la taille de la goutte déposée (Natrix39).

Cet essai m'a permis une première approche avec les techniques de cristallogénèse. Les échantillons de macromolécules utilisés sont des mélanges simples, dans les proportions d'une molécule d' $ARN_t$  pour deux molécules de SelA (ratio stoechiométrique). Les échantillons de protéines ecSelA et mtSelA( $\Delta 1-31$ ) utilisés ont été purifiés selon la méthode initiale de purification (voir § 3.3.3.1, page 145), et l' $ARN_t^{Sec}$  d'*E. coli* employé est issu du premier lot chargé en sérine.

L'apparition de cristaux se déroule, selon les conditions, entre 7 jours et 6 semaines après le dépôt des gouttes de cristallisation, puis ceux-ci croissent, de façon inversement proportionnelle à la quantité de points de nucléation observés.

Les conditions dans lesquelles les cristaux les plus intéressants sont apparus, sont testées dans un nouvel essai de cristallogénèse.

### Confirmation des premiers positifs

Les conditions paraissant les plus intéressantes sont Natrrix39 et Natrrix43, pour lesquelles un nombre modéré de cristaux de tailles appréciables sont apparus.

C'est donc avec ces conditions, toutes deux composées d'acétate d'ammonium, de chlorure de magnésium et d'un précipitant organique, à pH neutre, qu'une nouvelle cristallogénèse ciblée est tentée, un mois après le premier essai.

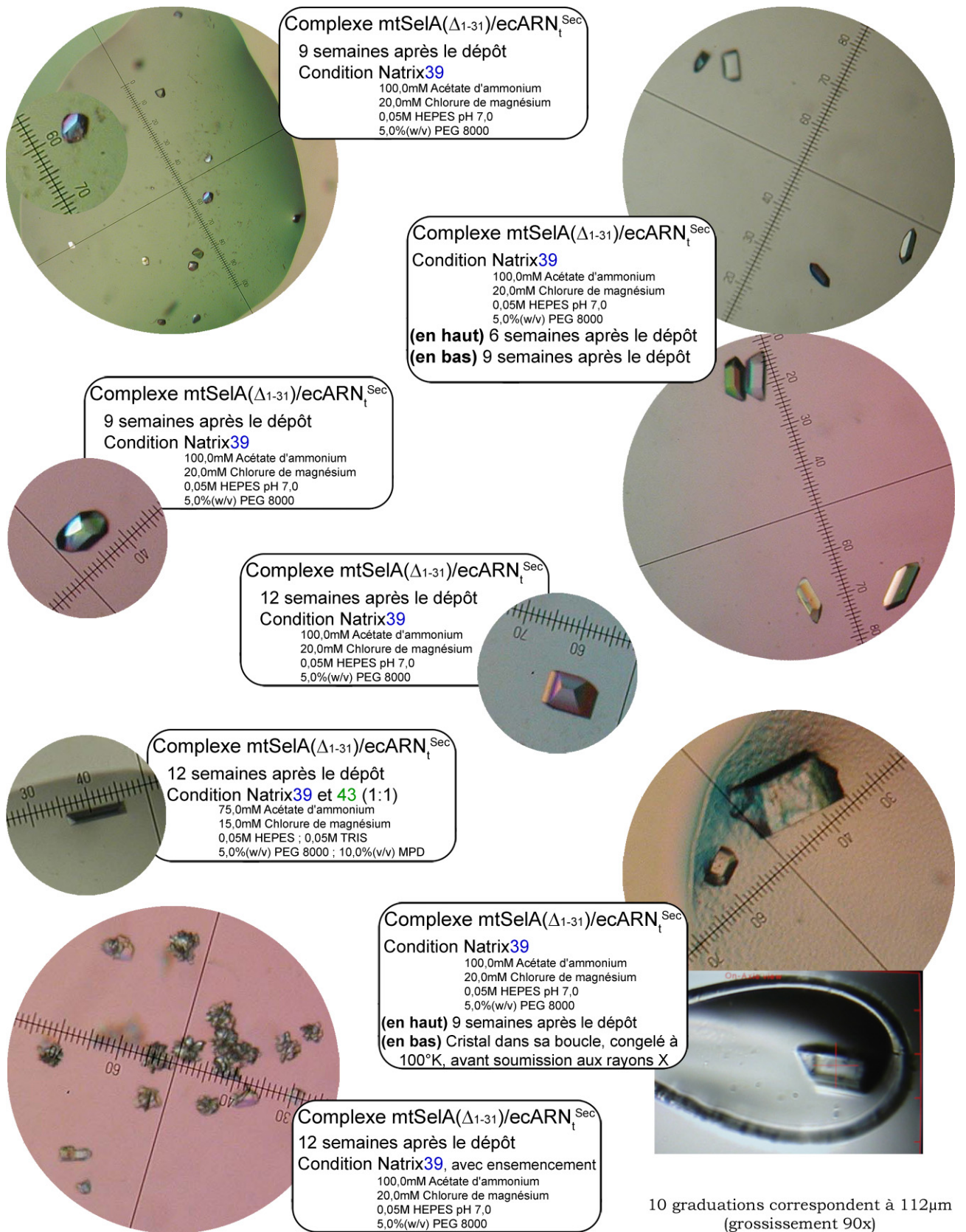
(25/11/2005)	Cristallogénèse n° 2
12 conditions	Variations autour des conditions Natrrix39 et 43
Gouttes suspendues, mise en place manuelle	
Espèces mises en oeuvre	Complexe mtSelA( $\Delta 1-31$ ) / $ARN_t^{Sec}$ d' <i>E. coli</i>
Formation du complexe	mélange simple $ARN_t^{Sec}$ / SelA (1:2)

**Tableau 3.5** – Résumé de la cristallogénèse n° 2.

Cet essai est d'autant plus important qu'un créneau de 48h a pu être réservé par notre équipe sur une ligne prévue pour la diffraction de cristaux de macromolécules aux rayons X, la ligne BM30 de l'ESRF (European Synchrotron Radiation Facility) à Grenoble, le 20 février 2006. Cela laisse donc 12 semaines aux potentiels cristaux pour acquérir une taille suffisante aux mesures de diffraction.

Les cristaux sont, là aussi, au rendez-vous, avec des morphologies très différentes par rapport à la cristallogénèse n°1 (voir les photos de la figure 3.17, page suivante).

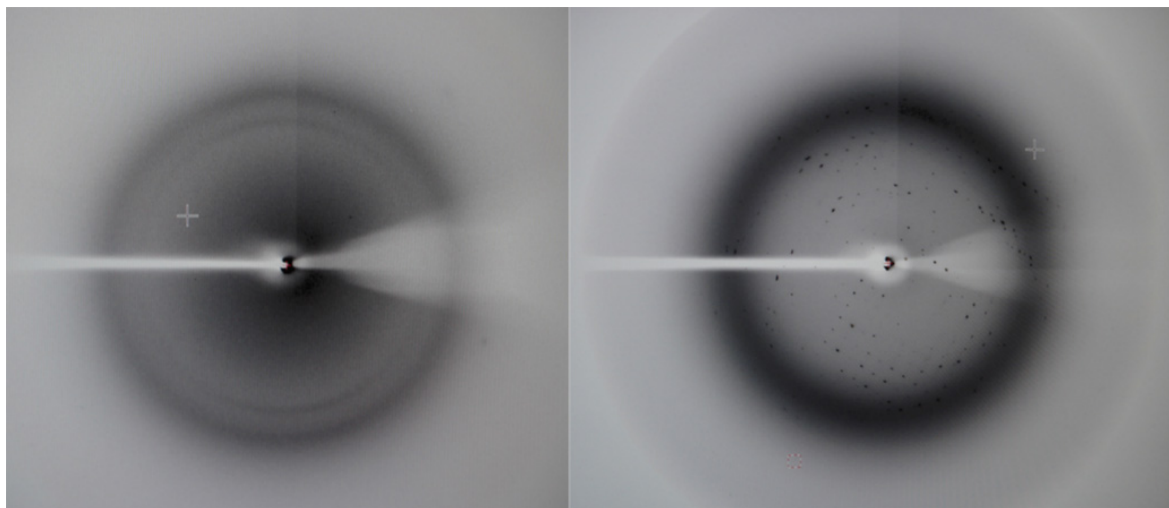
Les cristaux de tailles intéressantes ont été pêchés avec des micro-boucles prévues à cet effet, trempés dans une solution de cryoprotection, puis congelés dans de l'azote liquide. La solution de cryoprotection utilisée est un mélange de la condition de cristallisation et d'une substance ayant un effet chaotropique lors de la congélation : l'eau de la solution ne doit pas s'organiser en glace cristalline autour du cristal de la macromolécule, mais en une couche de glace amorphe n'interférant pas avec la diffraction aux rayons X.



**Figure 3.17** – Morceaux choisis de la Cristallogenèse n°2. Les couleurs des cristaux sont artificielles, dues à leur observation sous lumière polarisée. Une tentative d'ensemencement ("seeding") a été effectuée à partir des cristaux de la condition N39 de la cristallogenèse n°1, et les "oursins" de cristaux obtenus sont visibles en bas à gauche.

Dans notre cas, la condition de cristallisation est complétée par 30% de glycérol (v/v) pour obtenir la solution de cryoprotection, ou moins si elle contient déjà un composé organique, en tant que précipitant (MPD, PEG, isopropanol, *etc*).

Les résultats des mesures de diffraction aux rayons X furent tous négatifs, pour la trentaine de cristaux criblés, provenant des différentes conditions des essais de cristallogénèse n°1 et n°2.



**Figure 3.18** – À gauche, cliché de (non-)diffraction obtenu lors de la soumission aux rayons X du cristal représenté dans sa boucle sur la page précédente, comparé à un cliché de diffraction, à droite, obtenu pour un cristal de complexe de SelB complexée à SECIS, par Nicolas Soler.<sup>[11.23]</sup>

C'est devant ces mauvais résultats que j'ai décidé de modifier la méthode de purification de la protéine, car une des raisons pouvant expliquer l'absence de diffraction des cristaux obtenus est l'inhomogénéité de l'échantillon utilisé. Bien que cela soit généralement un frein à la cristallisation, il est possible que de tels échantillons donnent naissance à des empilements de macromolécules parsemés de défauts. Les mailles pseudo-cristallines ne sont alors pas constantes dans l'empilement obtenu, ce qui inhibe le phénomène de diffraction.

Le bilan de cette première campagne de cristallogénèse est tout de même positif, puisqu'elle m'a donnée la possibilité de découvrir les différentes étapes du processus, de l'obtention de cristaux à la mesure de leur diffraction aux rayons X. Les efforts à porter doivent être dirigés sur l'amélioration de chaque étape de ce processus complexe, destiné à l'obtention de la structure de notre complexe.

### Premier changement de protocole de purification

Peu après les essais de mesures de diffraction à l'ESRF, un premier essai de purification modifiée de mtSelA( $\Delta$ 1-31) fut basé sur l'enrichissement de la protéine par chromatographie répétée sur colonne d'affinité et sélection de la population de pureté la plus importante.

Cette technique permet de travailler dans un tampon unique, connu pour stabiliser la protéine et éviter sa précipitation.

(23/03/2006)	Cristallogenèse n° 3
98 conditions	kits Natrrix et Cristal Screen I de Hampton Research®
Gouttes suspendues, mise en place manuelle	
Espèces mises en oeuvre	Complexe mtSelA( $\Delta$ 1-31) / ARN <sub>t</sub> <sup>Sec</sup> d' <i>E. coli</i>
Formation du complexe	mélange simple ARN <sub>t</sub> <sup>Sec</sup> / SelA (1:2)

**Tableau 3.6** – Résumé de la cristallogenèse n° 3.

Cette tentative ne permit pas d'obtenir de cristaux, mais seulement des précipités de protéines ou des agrégats sphéroïdaux — ou sphérulites, quasi-cristaux — et fut donc abandonnée pour une nouvelle méthode.

Tous mes efforts se sont alors portés sur la mise au point d'une purification efficace, et sur la compréhension de la formation du complexe entre SelA et l'ARN<sub>t</sub><sup>Sec</sup>

### Evolution définitive du protocole de purification

De grandes mutations dans les techniques menant à l'obtention de cristaux, permettent d'aborder une nouvelle campagne de cristallisation sur de toutes nouvelles bases :

- L'utilisation d'une colonne de gel à exclusion stérique lors de la deuxième étape de purification de la protéine permet l'utilisation d'échantillons d'une excellente homogénéité.
- Les études menées sur le décimètre et la formation du complexe permettent une meilleure compréhension de l'interaction entre SelA et l'ARN<sub>t</sub><sup>Sec</sup>. Les nouveaux essais de cristallisation ont pour point de départ le complexe covalent purifié — produit à partir d'un mélange ARN<sub>t</sub><sup>Sec</sup> / SelA (3:1) — et ne se basent plus sur la simple présence de l'ARN<sub>t</sub><sup>Sec</sup> lors de la cristallisation — proportions stoechiométriques (1:2).
- L'accès à la plate-forme robotisée de cristallisation du LEBS permet un criblage étendu de conditions de cristallisation dans un temps relativement court : le faible volume de la goutte de cristallisation réduit la durée nécessaire au système pour atteindre l'équilibre par diffusion de vapeur. La durée moyenne d'apparition de cristaux passe, dans notre cas, de quelques semaines à quelques jours.

Le premier essai de cristallogenèse sur la plate-forme robotisée permit d'obtenir rapidement des cristaux, dans le cas du complexe ecSelA / ARN<sub>t</sub><sup>Sec</sup>. La protéine seule, elle, forme des précipités dans une majorité de condition de cristallisation, et cela dès la constitution de la goutte de mélange. Cela montre encore une fois le problème ardu que pose la solubilité de la protéine SelA.

(22/11/2006)	Cristallogenèse n° 4
288 conditions	kits Classics, Nucleix et PEGs I de Qiagen®
Gouttes assises, mise en place robotisée	
Espèces mises en oeuvre	Complexe ecSelA / ARN <sub>t</sub> <sup>Sec</sup> d' <i>E. coli</i> Protéine ecSelA seule
Formation du complexe	Complexe purifié, basé sur un mélange ARN <sub>t</sub> <sup>Sec</sup> /SelA (3:1)

**Tableau 3.7** – Résumé de la cristallogenèse n° 4.

Pour confirmer les conditions cristallogènes repérées grâce aux essais sur micro-plaques, un essai de cristallogenèse manuelle est effectué en gouttes suspendues, de volume plus important. Lors de cet essai, le complexe est utilisé à deux concentrations différentes, et le filtrat de concentration est également utilisé pour former des gouttes de cristallisation aux cotés de notre macromolécule étudiée (la concentration du complexe est effectuée par filtration sur membrane).

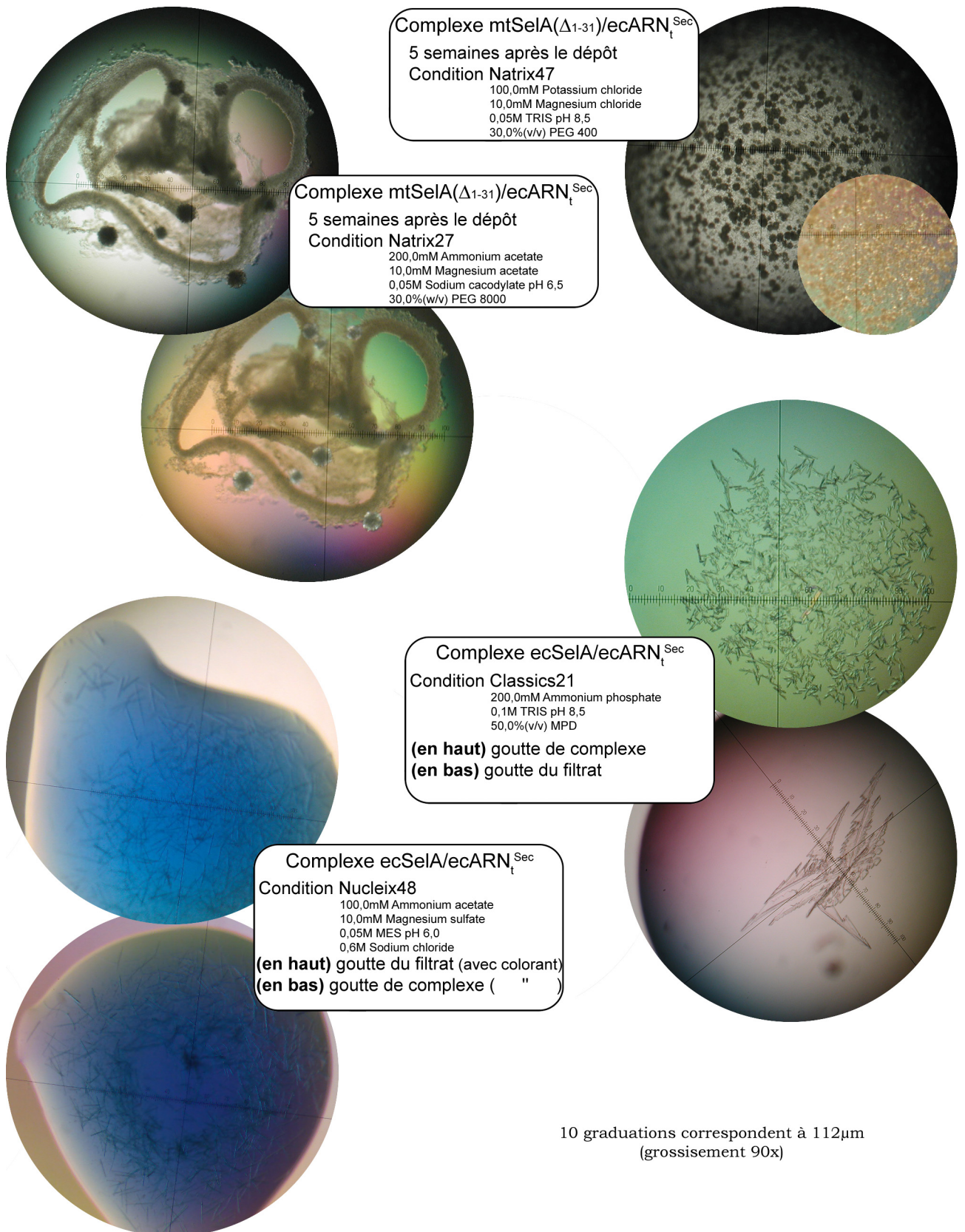
C'est grâce à cette précaution qu'il nous a été possible d'observer la formation de cristaux dans les gouttes de complexe, mais également dans certaines gouttes de filtrat, solution qui ne devrait contenir que le tampon utilisé lors de la purification du complexe.

(29/11/2006)	Cristallogenèse n° 5
20 conditions	extraites des kits Classics, Nucleix et PEGs I de Qiagen®
Gouttes suspendues, mise en place manuelle	
Espèces mises en oeuvre	Complexe ecSelA / ARN <sub>t</sub> <sup>Sec</sup> d' <i>E. coli</i>
Formation du complexe	Complexe purifié, basé sur un mélange ARN <sub>t</sub> <sup>Sec</sup> /SelA (3:1)

**Tableau 3.8** – Résumé de la cristallogenèse n° 5.

Le tampon de purification du complexe est justement un tampon simplifié contenant le minimum d'élément permettant la solubilisation de la protéine SelA (Tampon C : HEPES 50mM à pH 7,0 ; KCl 50mM ; imidazole 80mM ; DTT 1mM).

La vérification de la nature des cristaux obtenus dans les gouttes de filtrat et de complexe confirme qu'il s'agit de sels. Un colorant bleu spécifique est ajouté à la goutte contenant des cristaux. La petite molécule de colorant entre dans les canaux de solvant des cristaux de protéines, les colorant en bleu, alors que les cristaux de sels ne possèdent pas de canaux permettant cette coloration. À une dilution appropriée du colorant, les cristaux de protéines deviennent bleus dans une solution incolore, les cristaux de sels restent, eux, incolores, comme c'est le cas sur les photos de la figure 3.19, page ci-contre.



**Figure 3.19** – Morceaux choisis des cristallogenèses n°3 avec mtSelA( $\Delta_{1-31}$ ) et n°5 avec ecSelA. En haut, précipité et sphérulites semi-cristallines vus sous lumière classique et polarisée. En bas, cristaux salins des gouttes contenant du complexe et des gouttes contenant du filtrat de concentration.



Vue la composition du tampon de purification du complexe, les cristaux obtenus sont sans doute des cristaux de sels d'imidazole. La composition de ce tampon est donc modifiée, pour ne plus contenir d'imidazole (Tampon D : HEPES 20mM à pH 7,0 ; KCl 50mM ; DTT 1mM). Les tests de compatibilité ne montrent aucun changement dans le comportement du complexe, mais ce tampon ne convient plus à la solubilisation de la protéine seule, dont les essais de cristallisation sont abandonnés.

### Changement du tampon de cristallisation

Une nouvelle campagne de cristallisation est entamée, pour laquelle la purification de SelA est modifiée par l'ajout de l'étape de complémentation de la protéine en cofacteur PLP, et la purification du complexe par l'emploi du tampon sans imidazole.

(28/02/2007)	Cristallogénèse n° 6
288 conditions	kits Classics, Nucleix et PEGs I de Qiagen®
Gouttes assises, mise en place robotisée	
Espèces mises en oeuvre	Complexe ecSelA / ARN <sub>t</sub> <sup>Sec</sup> d' <i>E. coli</i> Complexe mtSelA / ARN <sub>t</sub> <sup>Sec</sup> d' <i>E. coli</i>
Formation du complexe	Complexe purifié, basé sur un mélange ARN <sub>t</sub> <sup>Sec</sup> /SelA (3:1)

**Tableau 3.9** – Résumé de la cristallogénèse n° 6.

Cette fois-ci, aucun cristal n'a pu être obtenu. Les concentrations des solutions de complexe étaient un peu inférieures à l'habitude, à cause de l'utilisation d'une quantité moins importante d'ARN<sub>t</sub><sup>Sec</sup>, ce qui peut expliquer ce résultat. Il a en plus été observé, lors de la purification des protéines SelA, une plus grande instabilité de ecSelA que de mtSelA.

J'ai donc décidé de concentrer les efforts de cristallogénèse sur les partenaires pouvant être produits dans les quantités les plus importantes, c-à-d. mtSelA et l'ARN<sub>t</sub><sup>Sec</sup> d'*E. coli*.

À cause de la complication croissante de la gestion des résultats de cristallogénèse, dans des conditions différentes et utilisant des kits de cristallisation de différentes marques, parfois redondants, j'ai mis en place une base de donnée compilant sous un seul format les compositions de tous les kits commerciaux auxquels nous avons accès. Cet outil permet une navigation facilitée dans l'espace des paramètres de cristallisation et une visualisation aisée, par un code couleur, des résultats des campagnes de cristallogénèse, dans le but d'identifier des conditions de cristallisation d'intérêt similaires ou ayant un paramètre en commun.

Cette base fut très utile lors de la campagne de cristallisation n°7, pour l'identification des tendances dans le comportement de cristallisation du complexe mtSelA / ecARN<sub>t</sub><sup>Sec</sup> à concentration élevée (15 mg/mL).

### 3.5.3 Cristaux et reproductibilité

Grâce aux apprentissages successifs apportés par les résultats des différentes cristallogenèses effectuées jusque là, la cristallogenèse n°7, réalisée avec une concentration importante de complexe de mtSelA et de ecARN<sub>t</sub><sup>Sec</sup> (15mg/mL) et dans des conditions strictement maîtrisées, a permis d'obtenir des cristaux de petite taille.

(05/04/2007)	Cristallogenèse n° 7
384 conditions	kits Classics, Nucleix et PEGs I & II de Qiagen®
Gouttes assises, mise en place robotisée	
Espèces mises en oeuvre	Complexe mtSelA / ARN <sub>t</sub> <sup>Sec</sup> d' <i>E. coli</i>
Formation du complexe	Complexe purifié, basé sur un mélange ARN <sub>t</sub> <sup>Sec</sup> /SelA (3:1)

**Tableau 3.10** – Résumé de la cristallogenèse n° 7.

Cette réussite montre bien la nécessité de l'utilisation de complexe issu du processus de production et de purification récemment mis au point, en grande quantité. Comme les points de nucléations furent nombreux, une concentration moins élevée fut utilisée par la suite.



**Figure 3.20** – Morceaux choisis de la Cristallogenèse n°7, points de départ d'une stratégie d'optimisation de conditions de cristallisation.

La reproduction des conditions de cristallisation ainsi déterminées a permis d'obtenir des cristaux intéressants, en micro-gouttes assises dans la condition Nucleix49 (cristallogénèse n°8). La même condition, en gouttes suspendues de volume plus important, n'a donné que des "oursins" ou des agrégats semi-cristallins de matériel protéique (voir les photos de la figure 3.21). Aucun cristal ne fut obtenu dans la condition Classics61.

(25/04/2007)	Cristallogénèse n° 8
Ensemble hybride de conditions Nucleix49 et Classics61	
96 conditions	Gouttes assises, mise en place robotisée
10 conditions	Gouttes suspendues, mise en place manuelle
Espèces mises en oeuvre	Complexe mtSelA / ARN <sub>t</sub> <sup>Sec</sup> d' <i>E. coli</i>
Formation du complexe	Complexe purifié, basé sur un mélange ARN <sub>t</sub> <sup>Sec</sup> /SelA (3:1)

**Tableau 3.11** – Résumé de la cristallogénèse n° 8.

J'entrepris donc une série d'essais de reproductibilité et d'amélioration des cristaux obtenus : j'ai d'abord pris soin de rester dans les mêmes conditions que celles qui ont permis d'obtenir les cristaux précédant, puis j'ai fait légèrement varier les paramètres de la solution de cristallisation d'intérêt (Nucleix49 : HEPES 50mM à pH 7,5 ; MgCl<sub>2</sub> ; Spermine 2,5mM), en modifiant le pH, la concentration des composants ou la présence en faible quantité d'additifs salins ou organiques.

(03/05/2007)	Cristallogénèse n° 9
12 conditions	Gradient de concentration de Nucleix49
Gouttes assises, mise en place manuelle	
Espèces mises en oeuvre	Complexe mtSelA / ARN <sub>t</sub> <sup>Sec</sup> d' <i>E. coli</i>
Formation du complexe	Complexe purifié, basé sur un mélange ARN <sub>t</sub> <sup>Sec</sup> /SelA (3:1)

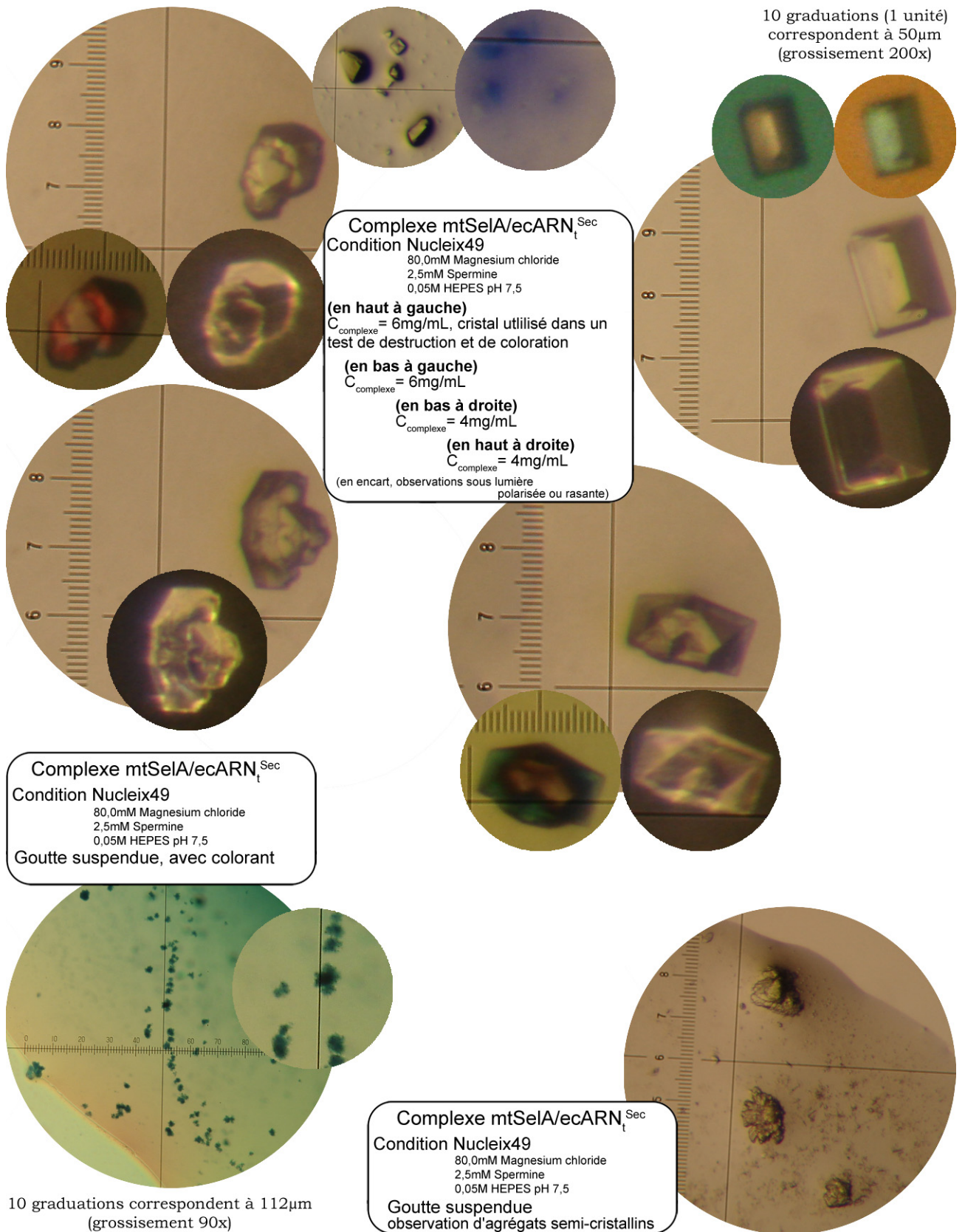
**Tableau 3.12** – Résumé de la cristallogénèse n° 9.

(02/08/2007)	Cristallogénèse n° 10
96 conditions	Variation du pH et apports d'additifs à Nucleix49
Gouttes assises, mise en place robotisée	
Espèces mises en oeuvre	Complexe mtSelA / ARN <sub>t</sub> <sup>Sec</sup> d' <i>E. coli</i>
Formation du complexe	Complexe purifié, basé sur un mélange ARN <sub>t</sub> <sup>Sec</sup> /SelA (3:1)

**Tableau 3.13** – Résumé de la cristallogénèse n° 10.

Malgré les efforts dans ce sens, je n'ai pu aboutir à une nouvelle formation de cristaux.

En entamant mon prolongement de financement de 6 mois, la décision fut prise de laisser en suspend le projet de cristallogénèse, pour me concentrer sur l'obtention de données biochimiques sur la formation du complexe entre SelA et l'ARN<sub>t</sub><sup>Sec</sup>, projet déjà entamé et mené en parallèle lors de la mise au point des conditions de purification des partenaires protéiques et nucléiques.



**Figure 3.21** – Cristallogenèse n°8 - essais d'optimisation de la cristallogenèse dans la condition Nucleix49, dans des micro-gouttes assises (en haut) et dans des gouttes suspendues (en bas).

Durant ce travail de cristallogenèse, des cristaux du complexe entre SelA et l'ARN<sub>t</sub><sup>Sec</sup> ont pu être obtenus, à plusieurs reprises. Ces résultats sont le fruit d'une optimisation des procédés de purification des partenaires, de formation du complexe et de sa purification, et enfin de la cristallogenèse en elle-même.

Malgré les efforts menés dans ce sens, la reproductibilité de la cristallogenèse reste un problème majeur de ce projet.

Les derniers cristaux intéressants obtenus (cristallogenèse n°8, photos de la page précédente) ont été montés sur des boucles adéquates pour la mesure de diffraction aux rayons X, transférés dans une solution cryoprotectante et congelés dans l'azote liquide.

Leur nombre restreint fut un frein pour la réservation de temps sur une ligne de diffraction aux rayons X. La vérification du potentiel de diffraction de ces cristaux est l'élément manquant pour valider le processus mis au point afin de les obtenir.

Néanmoins, une rapide analyse visuelle de ces cristaux permet de se rendre compte qu'ils sont maclés, c-à-d. que plusieurs arrangements cristallins cohabitent dans une même structure. Même en cas de diffraction de ces cristaux, ce qui représenterait déjà une avancée significative, les données obtenues, superposition de la diffraction de plusieurs réseaux cristallins, ne pourraient que très difficilement donner accès aux informations structurales sur le complexe.

Ces résultats sont une avancée significative dans l'obtention de cristaux du complexe protéo-nucléique entre SelA et l'ARN<sub>t</sub><sup>Sec</sup>, mais ne représentent qu'une étape sur la route de la résolution de la structure de ce complexe.

---

En marge de l'objectif de biologie structurale de ce projet de recherche, des analyses biochimiques sur le fonctionnement du complexe ont été effectuées, lors de la mise au point du processus aboutissant à la cristallogenèse.

Ces données accumulées ont été complétées, lors des derniers mois de mon travail de doctorat, par l'utilisation de nouvelles techniques, aboutissant à des conclusions intéressantes sur le comportement du complexe, présentées dans le chapitre suivant.

# Chapitre 4

## Etudes enzymatiques du couple SelA / ARN<sub>t</sub><sup>Sec</sup>

### Sommaire

---

<b>4.1</b>	<b>Formation du complexe</b>	<b>174</b>
4.1.1	Observation directe des produits de réaction	174
4.1.2	Synthèse chimique du sélénophosphate $\text{SePO}_3^{3-}$	176
<b>4.2</b>	<b>Mesures de protection à la déacylation</b>	<b>179</b>
4.2.1	Technique employée	180
4.2.2	Résultats obtenus	181
<b>4.3</b>	<b>Expériences d'empreintes chimiques</b>	<b>182</b>
4.3.1	La technique SHAPE	183
4.3.2	Empreintes sur l'ARN <sub>t</sub> <sup>Sec</sup> libre et complexé	184
4.3.3	Implications des résultats	186
<b>4.4</b>	<b>Cinétiques de la réaction de complexation</b>	<b>188</b>
4.4.1	Rappel de la problématique	188
4.4.2	Techniques de mesure	190
4.4.3	Mesures effectuées et résultats obtenus	192
4.4.4	Modélisation du comportement enzymatique de SelA	197

---

La recherche des meilleures conditions de purification des partenaires SelA et ARN<sub>t</sub><sup>Sec</sup> et de formation de leur complexe, nous a permis d'accéder à de nombreuses données sur leurs propriétés. Ce chapitre présente les expériences menées dans le but de connaître encore plus en détail le fonctionnement de l'enzyme SelA, le mode d'interaction de cette enzyme avec l'ARN<sub>t</sub><sup>Sec</sup> et les mécanismes régissant la formation du complexe entre ces deux partenaires.

## 4.1 Formation du complexe

Au début de mon doctorat, les techniques de chromatographie sur gel à exclusion stérique, permettant le suivi de la formation du complexe que nous souhaitions obtenir, n'étaient pas encore mises en place.

D'autres techniques ont donc été testées, pour vérifier que les protéines recombinantes produites interagissent effectivement avec le séryl-ARN<sub>t</sub><sup>Sec</sup> synthétisé par transcription *in vitro*. Il était en particulier nécessaire de s'assurer que l'étiquette portant les 6 histidines nécessaires à la purification des protéines sur colonne d'affinité, ne perturbe pas cette interaction, ce que nous avons pu vérifier ensuite par chromatographie.

### 4.1.1 Observation directe des produits de réaction

Une des méthodes de prédilection pour suivre la formation du complexe entre SelA et le séryl-ARN<sub>t</sub><sup>Sec</sup> est l'observation de l'apparition des produits de réaction.

Dans ce but, la technique la plus employée dans la littérature est l'utilisation de sérine<sup>14</sup>C lors de la charge de l'ARN<sub>t</sub><sup>Sec</sup>, et l'observation des produits radioactifs formés lors de la mise en présence de ce séryl<sup>14</sup>C-ARN<sub>t</sub><sup>Sec</sup> avec SelA.<sup>[II.8, II.9, II.14, II.28-II.31]</sup>

Ces produits sont observés grâce à la suite de manipulations suivantes : une déacylation en milieu basique de l'ARN<sub>t</sub><sup>Sec</sup>, puis une migration des produits de déacylation sur plaque de chromatographie sur couche mince (CCM), et enfin l'autoradiographie de la plaque de migration obtenue. Les positions de migration des produits radioactifs permettent d'en déduire la nature.

Les produits peuvent également être séparés, détectés et identifiés par utilisation d'un appareil HPLC, la radioactivité des fractions de séparation pouvant être mesurée par diverses techniques.<sup>[II.8]</sup>

Ces deux méthodes d'observation des produits réactionnels peuvent être employées lors d'une réaction de SelA avec le séryl-ARN<sub>t</sub><sup>Sec</sup> en présence ou en l'absence de sélénophosphate :

- Dans le premier cas, si la réaction a lieu, de la sélénocystéine<sup>14</sup>C est observée comme seul produit obtenu.

- Dans le deuxième cas, en l'absence de donneur de sélénium, la formation du complexe et sa décomposition produisent du pyruvate<sup>14</sup>C (voir le mécanisme réactionnel de la figure 2.7, page 95). L'intermédiaire réactionnel (2) peut également être réduit à l'aide d'un hydrure, comme NaBH<sub>4</sub>, et produire de l'alanine<sup>14</sup>C.

Dans le cas de la réaction complète de synthèse de la sélénocystéine en présence de sélénophosphate, ce dernier est généralement produit *in situ* grâce à la présence dans le milieu de réaction d'enzyme SelD purifiée, d'ATP et de sels de sélénures. Dans ce cas, l'élément radioactif utilisé lors de la réaction peut également être du sélénium  $^{75}\text{Se}$ .<sup>[II.14,II.29]</sup>

Nous avons voulu tester si la mise en place d'un tel protocole était possible avec le matériel disponible au laboratoire. De la sérine  $^{14}\text{C}$  était déjà utilisée lors de la mesure de la cinétique de charge de l' $\text{ARN}_t^{\text{Sec}}$ , c'est donc avec ce composé radioactif que nous avons effectué l'essai suivant, fin 2005.

Les réactions des protéines ecSelA et mtSelA( $\Delta 1-31$ ) sur du séryl  $^{14}\text{C}$ -  $\text{ARN}_t^{\text{Sec}}$  ont été effectuées en l'absence de sélénophosphate.

Des échantillons de séryl  $^{14}\text{C}$ - $\text{ARN}_t^{\text{Sec}}$  sont mis à réagir avec des quantités variables de SelA, dans un tampon à pH 7,3 (le "Pure system buffer"), pendant 15 minutes. Certains de ces échantillons sont additionnés d'hydrure  $\text{NaBH}_4$  et laissés à réagir 10 minutes supplémentaires.

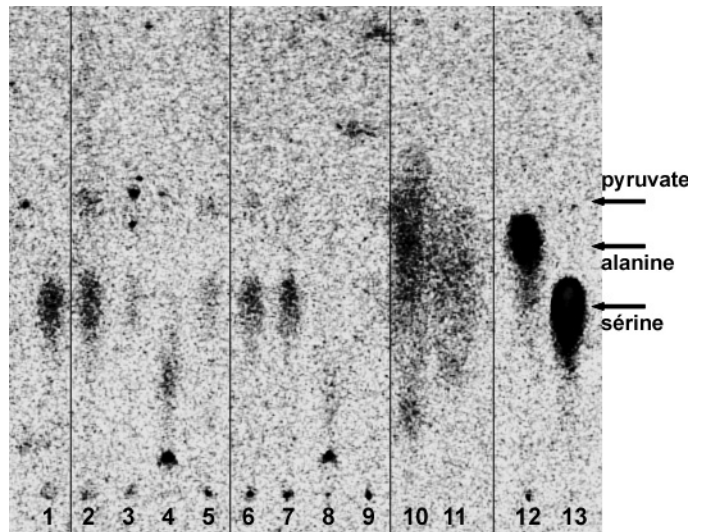
La réaction est arrêtée par précipitation des  $\text{ARN}_t^{\text{Sec}}$  à l'éthanol. Les précipités centrifugés sont lavés avec une solution d'éthanol à 70%, puis déacylés par redissolution et chauffage à  $40^\circ\text{C}$  pendant 40 minutes, dans un tampon de carbonate d'ammonium à pH 8,0. Après lyophilisation, les échantillons sont dissous dans un volume minimum d'eau pour leur dépôt sur la plaque CCM. Celle-ci est éluée dans un mélange eau/éthanol (3:7),<sup>[II.8]</sup> puis positionnée sur un écran sensible pendant une semaine. L'écran est ensuite "lu" par un appareil Storm Phosphorimager<sup>®</sup>, ce qui permet d'obtenir l'image 4.1, ci-dessus.

Le résultat obtenu n'est pas très satisfaisant.

Les différents dépôts sur la CCM correspondent aux échantillons suivants :

- ligne 1 : échantillon séryl  $^{14}\text{C}$ - $\text{ARN}_t^{\text{Sec}}$  de référence (6 pmol), sans protéine ;
- lignes 2 et 6 : réactions du séryl  $^{14}\text{C}$ - $\text{ARN}_t^{\text{Sec}}$  (6 pmol) avec 12 pmol de SelA ;
- lignes 3 et 7 : réactions avec 24 pmol de SelA ;
- lignes 4 et 8 : réactions avec 24 pmol de SelA, puis réduction par  $\text{NaBH}_4$  ;
- lignes 5 et 9 : réactions avec 24 pmol de SelA, puis dépôt sans déacylation ;
- lignes 10 et 11 : solutions d'éthanol de précipitation des échantillons 5 et 9, concentrées ;
- lignes 12 et 13 : échantillons d'alanine  $^{14}\text{C}$  et de sérine  $^{14}\text{C}$  de référence.

Différents profils de migration montrent la présence de pyruvate  $^{14}\text{C}$ , produit de réaction



**Figure 4.1** – Autoradiographie de la CCM de migration des produits de la réaction entre du séryl  $^{14}\text{C}$ - $\text{ARN}_t^{\text{Sec}}$  et mtSelA( $\Delta 1-31$ ) (lignes 2 à 5) ou ecSelA (lignes 6 à 9).



du séryl<sup>14</sup>C-ARN<sub>t</sub><sup>Sec</sup> avec ecSelA et mtSelA(Δ1-31), comme pour les échantillons 2, 5, 6, 7 et 9, en très faible quantité. Par contre, les profils de migration des échantillons réduits 4 et 8 ne montrent pas de trace d'alanine<sup>14</sup>C.

Le dépôt des échantillons non-déacylés 5 et 9, et de leurs solutions de précipitation en 10 et 11, avait pour but de vérifier la répartition de la radioactivité entre l'ARN<sub>t</sub><sup>Sec</sup> précipité, et la solution de précipitation contenant les possibles acides aminés saponifiés durant la réaction. Il est visible qu'une grande quantité de la radioactivité mise en jeu se retrouve finalement dans la solution de précipitation, non fixée à l'ARN<sub>t</sub><sup>Sec</sup>.

De plus, la quantité de radioactivité totale des échantillons 5 et 9 devraient se trouver au niveau de leurs points de dépôt, soit, vu les quantités de matériel utilisées, un cinquième de la radioactivité de l'échantillon 13 de sérine<sup>14</sup>C de référence. Au lieu de cela, non seulement leur activité est très faible, mais on observe de la sérine<sup>14</sup>C et du pyruvate<sup>14</sup>C sur le profil de migration de ces échantillons, ce qui indique la poursuite de la déacylation de l'ARN<sub>t</sub><sup>Sec</sup>.

Cet essai permet bien d'aboutir à la conclusion de la formation des complexes entre SelA et le séryl-ARN<sub>t</sub><sup>Sec</sup>, puisque du pyruvate<sup>14</sup>C est visible sur la CCM. Mais, vu la faiblesse de la radioactivité observée, cette conclusion demande vérification par une autre technique, ou l'amélioration de celle-ci.

Le problème de cet essai d'observation des produits de réaction est la déacylation de l'ARN<sub>t</sub><sup>Sec</sup> : le pH auquel est menée la réaction doit favoriser cette déacylation, ce qui explique la faible quantité de radioactivité présente dans les échantillons précipités, et la présence de radioactivité dans la solution d'éthanol de précipitation (lignes 10 et 11).

Comme les expériences de cinétiques de déacylation en fonction du pH n'avaient pas encore été effectuées à cette époque (voir § 4.2, page 179), la légère basicité du tampon de réaction n'avait pas été à priori la première des raisons mises en cause dans le résultat mitigé obtenu. Le problème d'illisibilité des résultats peut également être en partie la conséquence de l'utilisation du matériel de révélation disponible au laboratoire, dédié à l'utilisation de phosphore <sup>32</sup>P à l'activité spécifique beaucoup plus importante que celle du carbone <sup>14</sup>C.

Par la suite, la mise au point de l'utilisation d'une colonne de filtration sur gel à exclusion stérique a permis la confirmation de la formation du complexe entre les protéines recombinantes SelA et le séryl-ARN<sub>t</sub><sup>Sec</sup> produit par transcription *in vitro*.

#### 4.1.2 Synthèse chimique du séléno phosphate SePO<sub>3</sub><sup>3-</sup>

Comme je le rappelais dans la première partie de ce manuscrit, l'étude de la machinerie de biosynthèse et d'insertion de la séléno cystéine chez les eubactéries est un domaine de choix pour chercher les réponses aux interrogations laissées par cet acide aminé, et mieux comprendre, par cette exception, le fonctionnement et l'émergence du code génétique et de la traduction ribosomale.

Les deux directions thématiques de recherche abordées par l'équipe sont, d'un coté, l'étude des acteurs de l'incorporation de la sélénocystéine lors de la traduction, c-à-d. SelB interagissant, entre autres, avec son partenaire SECIS de l'ARN<sub>m</sub>, sujet de thèse<sup>[II.23]</sup> et de stage post-doctoral de Nicolas Soler, et de l'autre coté, l'étude des acteurs de la biosynthèse de la sélénocystéine, c-à-d. SelA et son interaction avec ARN<sub>t</sub><sup>Sec</sup>, sujet de ce présent travail.

Un des grands intérêts de travailler pour un grand projet thématique comprenant plusieurs directions de recherche parallèles, comme dans notre laboratoire, est qu'à certains moments, ces directions parallèles se croisent et interagissent pour bénéficier des avancées l'une de l'autre.

Comme dans le mécanisme de synthèse des sélénoprotéines dans lequel l'ARN<sub>t</sub><sup>Sec</sup>, présent à chaque étape, est l'élément pivot, c'est autour de ce même ARN<sub>t</sub><sup>Sec</sup> qu'interagissent mon travail et celui de Nicolas. Je dois ainsi mon apprentissage des techniques de synthèse du séryl-ARN<sub>t</sub><sup>Sec</sup>, au début de mon doctorat, à son expérience de la question. Aujourd'hui, les avancées dans la compréhension de la réaction du séryl-ARN<sub>t</sub><sup>Sec</sup> avec SelA, effectuées durant mon doctorat, nous permettent d'envisager la production de sélénocystéinyl-ARN<sub>t</sub><sup>Sec</sup>. Ce produit de la réaction complète du séryl-ARN<sub>t</sub><sup>Sec</sup> avec SelA en présence de sélénophosphate est un des partenaire de SelB, et sa production permettrait l'étude des interactions tripartites entre SelB, l'ARN<sub>t</sub><sup>Sec</sup> et le SECIS de l'ARN<sub>m</sub>, prolongement du sujet originel de Nicolas, ce qui est d'un grand intérêt dans la thématique de recherche du laboratoire.

Cette synthèse permet également de s'assurer de la réalité de l'activité sélénocystéine synthase des protéines recombinantes produites au cours de mon travail de doctorat.

Le sélénophosphate (voir § 2.2, page 87) peut être produit chimiquement, ou enzymatiquement grâce à l'action de SelD sur l'ATP et le sélénium minéral réduit sous forme de séléniures. Pour des raisons de commodités, les laboratoires de biologie préfèrent souvent cette seconde méthode à la première, qui est pourtant celle que nous avons suivie ici.

Deux synthèses chimiques du sélénophosphate sont décrites dans la littérature :

- Une première méthode passe par la réaction de la tris(triméthylsilyl)-phosphite (TMSO)<sub>3</sub>P sur du sélénium élémentaire Se<sup>0</sup>, puis l'hydrolyse du tris(triméthylsilyl)-sélénophosphate (TMSO)<sub>3</sub>PSe obtenu dans un tampon aqueux.<sup>[II.32, II.33]</sup>
- La deuxième méthode est décrite comme plus pratique que la précédente,<sup>[II.34, II.35]</sup> et c'est donc celle que j'ai employée ici. Elle passe par la réaction de la triméthyl-phosphite (CH<sub>3</sub>O)<sub>3</sub>P sur du sélénium élémentaire Se<sup>0</sup> pour obtenir le triméthyl-sélénophosphate (CH<sub>3</sub>O)<sub>3</sub>PSe,<sup>[II.36]</sup> puis par sa réaction avec de l'iodure de triméthylsilyl pour former le tris(triméthylsilyl)-sélénophosphate (TMSO)<sub>3</sub>PSe. L'hydrolyse de ce composé aboutit au sélénophosphate. (voir le schéma 4.2, page suivante.)

La réalisation de cette synthèse demande l'utilisation d'installations et de matériel classiques dans un laboratoire de chimie, absents de notre laboratoire de biologie moléculaire. Je remercie donc encore le Docteur Angela Marinetti pour m'avoir ouvert les portes



Cette réaction a permis d'obtenir une solution de tris(triméthylsilyl)-sélénophosphate  $(\text{TMSO})_3\text{PSe}$  à 120mM dans le chloroforme, conservée en aliquotes de 1mL sous atmosphère d'azote et pouvant être conservée à  $-20^\circ\text{C}$  ou à  $-80^\circ\text{C}$  pendant plus d'un an. Ces aliquotes permettent de disposer de façon immédiate, par hydrolyse sous atmosphère inerte et dans un tampon dégazé approprié, de sélénophosphate en quantité importante pour effectuer les réactions de synthèse de la sélénocystéine.

La mise au point de la réaction de contrôle de l'activité sélénocystéine synthase des protéines SelA recombinantes produites lors de mon doctorat a été entamée peu de temps avant le début de la rédaction de ce manuscrit. La technique employée est celle de l'autoradiographie d'une CCM de migration des produits radioactifs de la réaction, déjà utilisée au début de mon projet. La connaissance des problématiques qui peuvent influencer la lisibilité des résultats, permet d'envisager un meilleur déroulement lors de la nouvelle utilisation de cette technique.

Ainsi, après cette vérification, la synthèse de sélénocystéyl- $\text{ARN}_t^{\text{Sec}}$  en atmosphère inerte, pour l'étude de son interaction avec SelB, pourrait être possible.

## 4.2 Mesures de protection à la déacylation

Au cours de la mise au point du processus de purification des protéines SelA recombinantes, des mesures de stabilité de ces protéines ont été effectuées (voir § 3.3.3.3, page 147), et ont permis de conclure, entre autre, à une meilleure stabilité de ces protéines à pH basique. Or nous savons que le séryl- $\text{ARN}_t^{\text{Sec}}$  est sensible aux milieux basiques et se déacyle plus facilement que les autres aminoacyl- $\text{ARN}_t$ .

Nous avons donc effectué une mesure de la cinétique de déacylation du séryl- $\text{ARN}_t^{\text{Sec}}$  dans plusieurs tampons à différents pH. Nous avons également observé si l'ajout des protéines recombinantes SelA dans le milieu modifiait cette cinétique. Si une modification de la cinétique est observable, cela permettrait de conclure que la formation du complexe a bien lieu, et que la liaison ester entre l'acide aminé et l' $\text{ARN}_t^{\text{Sec}}$  est en partie protégée dans cette formation protéo-nucléique.

Cette technique est, en effet, utilisée dans différents travaux<sup>[II.30,II.37]</sup> pour montrer s'il existe une complexation de différents types d' aminoacyl- $\text{ARN}_t^{\text{Sec}}$  sauvages ou modifiés avec SelB, ou avec le facteur d'élongation EF-Tu.

Les résultats obtenus prouvent l'interaction de SelB d'*E. coli* avec le sélénocystéinyl- $\text{ARN}_t^{\text{Sec}}$  d'*E. coli* sauvage, sa version modifiée avec une tige-boucle variable d' $\text{ARN}_t^{\text{Ser}}$ , ou encore le sélénocystéinyl- $\text{ARN}_t^{\text{Sec}}$  humain.

Ils montrent également l'absence d'interaction entre SelB et le séryl- $\text{ARN}_t^{\text{Sec}}$  ou la version du sélénocystéinyl- $\text{ARN}_t^{\text{Sec}}$  portant une tige acceptrice raccourcie à 7 paires de bases, et l'absence d'interaction entre EF-Tu et le sélénocystéinyl- $\text{ARN}_t^{\text{Sec}}$  sauvage. (voir § 2.1.1, page 82, sur les déterminants de l' $\text{ARN}_t^{\text{Sec}}$ )

L'utilisation de cette technique pour caractériser la complexation du séryl- $ARN_t^{Sec}$  par SelA n'a pas été décrite.

### 4.2.1 Technique employée

La mesure de la cinétique de déacylation suit un protocole similaire à la mesure de la cinétique de charge en sérine $^{14}C$  de l' $ARN_t^{Sec}$  par la SerRS (voir § 3.2.3, page 140).

Une solution de tampon à un pH donné contenant une quantité connue de séryl $^{14}C$ - $ARN_t^{Sec}$  est mise à incuber à 37°C. Des prélèvements d'un volume précis de milieu réactionnel sont effectués à des temps donnés, et déposés sur du papier Whatman®. Ce papier est immédiatement plongé dans un bain d'acide trichloroacétique, permettant la précipitation des  $ARN_t^{Sec}$  sur le papier et le lavage de la sérine $^{14}C$  libérée par la déacylation éventuelle du séryl $^{14}C$ - $ARN_t^{Sec}$ . La quantité de radioactivité présente sur le papier, représentant la sérine $^{14}C$  toujours fixée à l' $ARN_t^{Sec}$ , est comptée par scintimétrie et comparée à la quantité de radioactivité qui devrait être théoriquement présente dans le volume prélevé, fonction de la quantité initiale de séryl $^{14}C$ - $ARN_t^{Sec}$  présente dans l'échantillon.

Les tampons pour lesquels la manipulation a été effectuée sont :

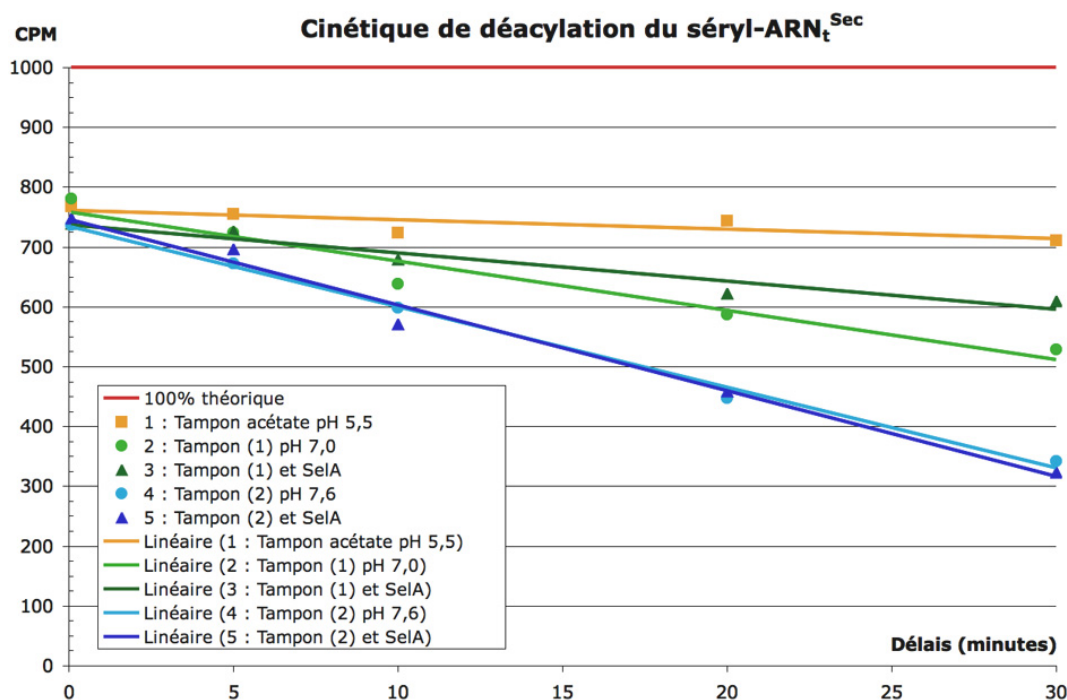
- un tampon acétate de sodium 100mM à pH 5,5 (Référence) ;
- un tampon HEPES-KOH 100mM à pH 7,0 (Tampon 1), contenant 80mM d'imidazole, 20mM de KCl, 10mM de  $MgCl_2$  et 1mM de DTT ;
- un tampon HEPES-KOH 100mM à pH 7,6 (Tampon 2), contenant les mêmes composants dans les mêmes proportions.

En plus de la simple mesure de la cinétique de déacylation du séryl $^{14}C$ - $ARN_t^{Sec}$  dans ces tampons, un deuxième lot d'échantillons est préparé pour observer l'effet de la présence de la protéine recombinante ecSelA sur cette cinétique, dans les tampons neutre et basique ci-dessus.

La mise en place des échantillons suit un protocole spécifique. La solution tampon est préparée et complétée avec la protéine ecSelA concentrée le cas échéant : le faible volume de tampon HEPES 50mM à pH 7,6 importé avec la protéine est compensé par l'ajout de la même quantité de tampon, sans protéine, dans les échantillons dans lesquels la protéine ecSelA est absente. Une mesure du changement de pH causé par l'ajout de ce faible volume de tampon dans la solution à pH 7,0 a révélé un effet minime sur celui-ci.

Une fois ces différentes solutions à 37°C, le séryl $^{14}C$ - $ARN_t^{Sec}$  y est ajouté, puis après homogénéisation, le premier prélèvement est effectué, soit 5 secondes après la mise en condition de l' $ARN_t^{Sec}$ . Ce premier prélèvement correspond au temps "0" des résultats, présentés sur le graphique de la figure 4.4, ci-après.

## 4.2.2 Résultats obtenus



**Figure 4.4** – Évolution de la quantité de sérine<sup>14</sup>C fixée à l'ARN<sub>t</sub><sup>Sec</sup>, en fonction du pH de la solution et de la présence ou non de protéine ecSelA.

La raison pour laquelle la quantité de radioactivité mesurée dans l'échantillon à  $t=0$  est inférieure à la quantité théorique, est que le calcul de la quantité théorique de CPM ne tient pas compte du rendement de charge de la réaction de production du séryl<sup>14</sup>C-ARN<sub>t</sub><sup>Sec</sup>.

Cette réaction de charge a donc un rendement d'environ 75%, à l'erreur de mesure près, ce qui correspond au rendement moyen de cette réaction pour le protocole de charge alors employé (voir § 3.2.3, page 140).

Comme on peut s'y attendre, la vitesse de déacylation du séryl<sup>14</sup>C-ARN<sub>t</sub><sup>Sec</sup> est une fonction croissante du pH de la solution. Cet ARN<sub>t</sub> se déacyle spontanément à pH 7,0, même si cette dégradation reste assez lente.

Cela confirme qu'il est nécessaire de conserver l'ARN<sub>t</sub><sup>Sec</sup> chargé en sérine dans des conditions acides.

Les résultats de cette expérience, lorsque la protéine ecSelA est présente dans le milieu, sont plus difficiles à interpréter. Alors qu'à pH 7,0, la vitesse de déacylation du séryl<sup>14</sup>C-ARN<sub>t</sub><sup>Sec</sup> semble diminuée avec la présence de ecSelA, la protéine ne semble avoir aucun effet sur la déacylation lorsque le pH est de 7,6.

Le résultat obtenu pour un pH neutre permet de confirmer la formation du complexe entre *ecSelA* et le séryl- $ARN_t^{Sec}$  et d'observer une protection de la liaison ester liant l'acide aminé à l' $ARN_t^{Sec}$  dans le complexe, à ce pH.

Plusieurs hypothèses permettent d'interpréter le résultat obtenu à pH 7,6 : soit ces conditions de pH font changer la polarité de certains sites de l'enzyme *ecSelA* chargés de l'interaction avec l' $ARN_t^{Sec}$ , ce qui perturbe la formation du complexe à ce pH, soit le complexe se forme tout de même à ce pH, mais la liaison ester liant l'acide aminé à l' $ARN_t^{Sec}$  est toujours accessible au solvant, et la déacylation s'effectue toujours au même rythme qu'en l'absence de *SelA*. Nous n'avons pas les éléments pour trancher entre ces deux hypothèses.

Ces données permettent de tirer une conclusion importante pour notre étude : pour des raisons de compatibilité entre les conditions de stabilité de la protéine seule et du séryl- $ARN_t^{Sec}$ , la formation du complexe, en vue de l'étude de ses propriétés ou de sa cristallisation, ne peut se faire qu'à pH neutre.

### 4.3 Expériences d'empreintes chimiques

L'étude de la structure secondaire des ARN est facilitée par l'existence d'outils adaptés, développés pour cette usage. En particulier, des expériences d'empreintes sur l'ARN par attaques de réactifs chimiques ou enzymatiques, ou par clivage du squelette ribophosphate, permettent de déterminer quelles régions de l'ARN est engagée dans des interactions structurantes et protectrices vis-à-vis de ces attaques, et quelles autres régions sont accessibles à ces réactifs.<sup>[II.38]</sup>

Généralement, les réactifs employés réagissent chacun sélectivement avec une certaine fonction chimique d'un type de base nucléique, ou avec certains sites du squelette de l'ARN, et il est donc nécessaire de combiner les résultats d'attaques de plusieurs réactifs sur l'ARN pour arriver à déterminer de façon globale quelle région est réactive et quelle région est protégée.

Certaines de ces techniques d'empreintes sont également utilisées pour la caractérisation de complexes protéo-nucléiques, et permettent en particulier de déterminer avec précision les éléments d'interactions spécifiques de l'ARN avec la protéine en question, par différence entre le profil de réactivité de l'ARN libre et de l'ARN en interaction avec son partenaire.

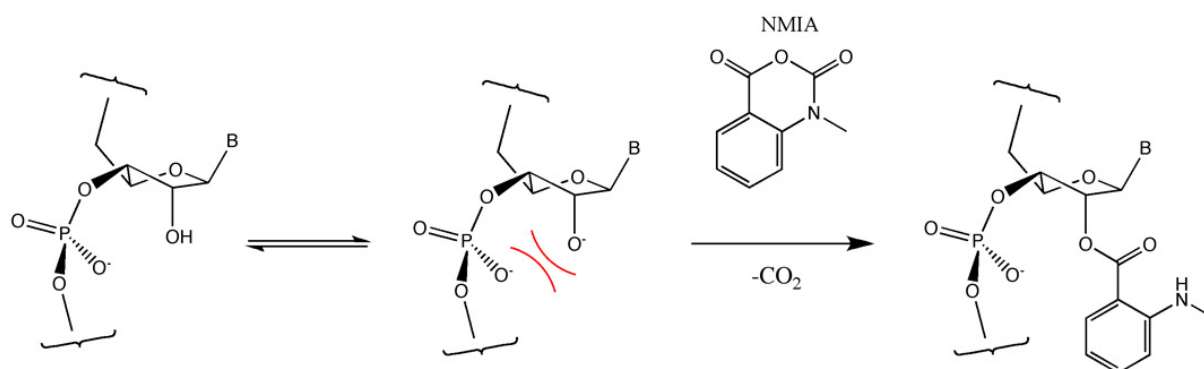
Récemment, une nouvelle technique d'empreinte chimique a été développée<sup>[II.39–II.42]</sup> qui permet une mesure locale de la flexibilité du squelette ribo-phosphate, nucléotide par nucléotide. Elle permet donc de savoir si un nucléotide est mobile ou fixé car impliqué dans une interaction, par appariement avec un nucléotide complémentaire ou par implication dans une structure tertiaire de l'ARN.

Cette technique a été baptisée SHAPE pour "Selective 2'-Hydroxyl Acylation analyzed by Primer Extension".

### 4.3.1 La technique SHAPE

Cette technique exploite la différence de réactivité de l'hydroxyle 2'-OH du ribose du squelette de l'ARN, en fonction de son environnement. Cette réactivité est en particulier modulée par la proximité de l'anion phosphodiester adjacent, présent en 3' du ribose.

L'acylation sélective de cette position 2' par l'anhydride N-méthylisatoïque (NMIA) demande le passage par un intermédiaire réactif : d'après Merino *et al.*,<sup>[II.39]</sup> la réaction ne se produit que lorsque l'hydroxyle 2'-OH du ribose est dans sa forme réactive minoritaire déprotonée, sous laquelle l'attaque nucléophile de l'anhydride conduit à la formation de l'ester (voir la figure 4.5). Or la formation de ce 2'-oxyanion est défavorisée lorsque le nucléotide se trouve dans une conformation contrainte entraînant une trop grande proximité entre l'hydroxyle 2'-OH et le phosphodiester chargé négativement présent en 3'.



**Figure 4.5** – Mécanisme d'acylation de l'hydroxyle 2'-OH du ribose d'un nucléotide flexible par l'anhydride N-méthylisatoïque, d'après Merino *et al.*<sup>[II.39]</sup>

Il a pu être montré que l'acylation sélective de cette position 2'-OH du ribose par l'anhydride N-méthylisatoïque était uniquement dépendante de la flexibilité locale du nucléotide, et donc de son implication dans les structures secondaires ou tertiaires de l'ARN considéré.<sup>[II.39, II.40]</sup>

Un facteur 10 sépare en effet la réactivité d'un nucléotide présent sur un simple brin d'ARN, d'un nucléotide engagé dans un appariement double brin, et la réaction ne semble pas dépendante de l'accessibilité de l'hydroxyle 2'-OH au solvant.<sup>[II.39]</sup>

Pour accéder aux résultats de la réaction d'acylation, une transcription inverse (ou rétrotranscription) est effectuée sur l'ARN à l'aide d'amorces adéquates, fonctionnalisées par un groupement fluorescent. La présence de l'ester volumineux en position 2' provoque l'arrêt de la rétrotranscriptase, après l'incorporation dans le brin rétrotranscrit du complémentaire au nucléotide précédant celui qui a été modifié. Ces rétrotranscrits abortifs sont séparés selon leurs tailles par électrophorèse sur gel, où leurs positions sont détectées par fluorescence.



Cette technique a permis à l'équipe qui l'a développée d'analyser la flexibilité locale de plusieurs ARN biologiques, comme l'ARN<sub>t</sub><sup>Asp</sup> de la levure *Saccharomyces cerevisiae*<sup>[II.39]</sup> ou un domaine de la RNase P de *Bacillus subtilis*.<sup>[II.42]</sup>

### 4.3.2 Empreintes sur l'ARN<sub>t</sub><sup>Sec</sup> libre et complexé

La technique SHAPE peut nous permettre d'accéder aux informations de flexibilité locale de l'ARN<sub>t</sub><sup>Sec</sup>, et de les comparer à celles observées pour l'ARN<sub>t</sub><sup>Asp</sup>.<sup>[II.39]</sup>

L'utilisation de cette technique d'empreintes chimiques sur un complexe protéo-nucléique n'a pas encore été décrite. Pourtant il paraît intéressant de mesurer l'influence de l'interaction entre une protéine et un ARN partenaire sur la flexibilité locale de ce dernier : c'est ce que nous avons fait pour le complexe Sela / séryl-ARN<sub>t</sub><sup>Sec</sup>.

La figure 4.6, page suivante, montre le résultat d'un essai de la technique SHAPE sur l'ARN<sub>t</sub><sup>Sec</sup> seul et sur cet ARN présent dans le complexe Sela / séryl-ARN<sub>t</sub><sup>Sec</sup>. Le complexe utilisé pour cette manipulation a été fraîchement purifié par la technique de chromatographie sur gel d'exclusion stérique : la population d'ARN<sub>t</sub><sup>Sec</sup> présente dans l'échantillon utilisé est totalement liée à la protéine, sous la forme du complexe intermédiaire (2) (voir le schéma réactionnel 2.7, page 95).

Sur l'image de gel d'électrophorèse observé par fluorescence, figure 4.6, le premier profil de migration correspond à l'échantillon témoin, c-à-d. à la rétrotranscription d'un échantillon d'ARN<sub>t</sub><sup>Sec</sup> qui n'a pas été soumis à la modification chimique par le NMIA.

La deuxième colonne correspond, elle, aux rétrotranscrits obtenus à partir d'ARN<sub>t</sub><sup>Sec</sup> libres soumis au réactif NMIA, et la colonne suivante, au résultat de rétrotranscription de l'ARN<sub>t</sub><sup>Sec</sup> présent dans le complexe, lui aussi soumis à la réaction d'empreintes chimiques.

Le séquençage de l'ARN<sub>t</sub><sup>Sec</sup> est, lui, effectué par rétrotranscription de quatre échantillons natifs en présence d'un des quatre 2',3'-didéoxy-nucléotides triphosphates, et est utilisé pour déterminer la taille des rétrotranscrits obtenus lors de l'expérience d'empreintes. Le séquençage obtenu est indiqué sur la droite de l'image du gel, avec le code couleur suivant désignant les nucléotides d'arrêt de la rétrotranscription pour les expériences d'empreintes : en rouge sont indiqués les nucléotides appartenant à la boucle anticodon de l'ARN<sub>t</sub><sup>Sec</sup>, correspondant à la zone notée (1) ; en bleu ceux appartenant à la tige-boucle variable, correspondant à la zone notée (2) ; et en vert tous les autres nucléotides pour lesquels un arrêt de la rétrotranscription est observable.

Sur le schéma de la structure secondaire de l'ARN<sub>t</sub><sup>Sec</sup> d'*E. coli* présent à droite de la figure, sont marqués en jaunes, d'une part les nucléotides d'hybridation de l'amorce (primer) de rétrotranscription, et d'autre part ceux pour lesquels la longueur du brin rétrotranscrit n'est pas suffisante pour qu'ils soient visibles sur le gel. Les bandes correspondant au rétrotranscrit complet, indiqué en gris au niveau du gel, masquent les informations concernant les deux nucléotides précédents G2 et A3.

Les nucléotides qui ont été modifiés par la réaction d'empreinte sont indiqués en mauve sur le schéma de structure secondaire de l'ARN<sub>t</sub><sup>Sec</sup> d'*E. coli* : il s'agit des nucléotides suivant ceux identifiés grâce à la taille des brins rétrotranscrits sur le gel.

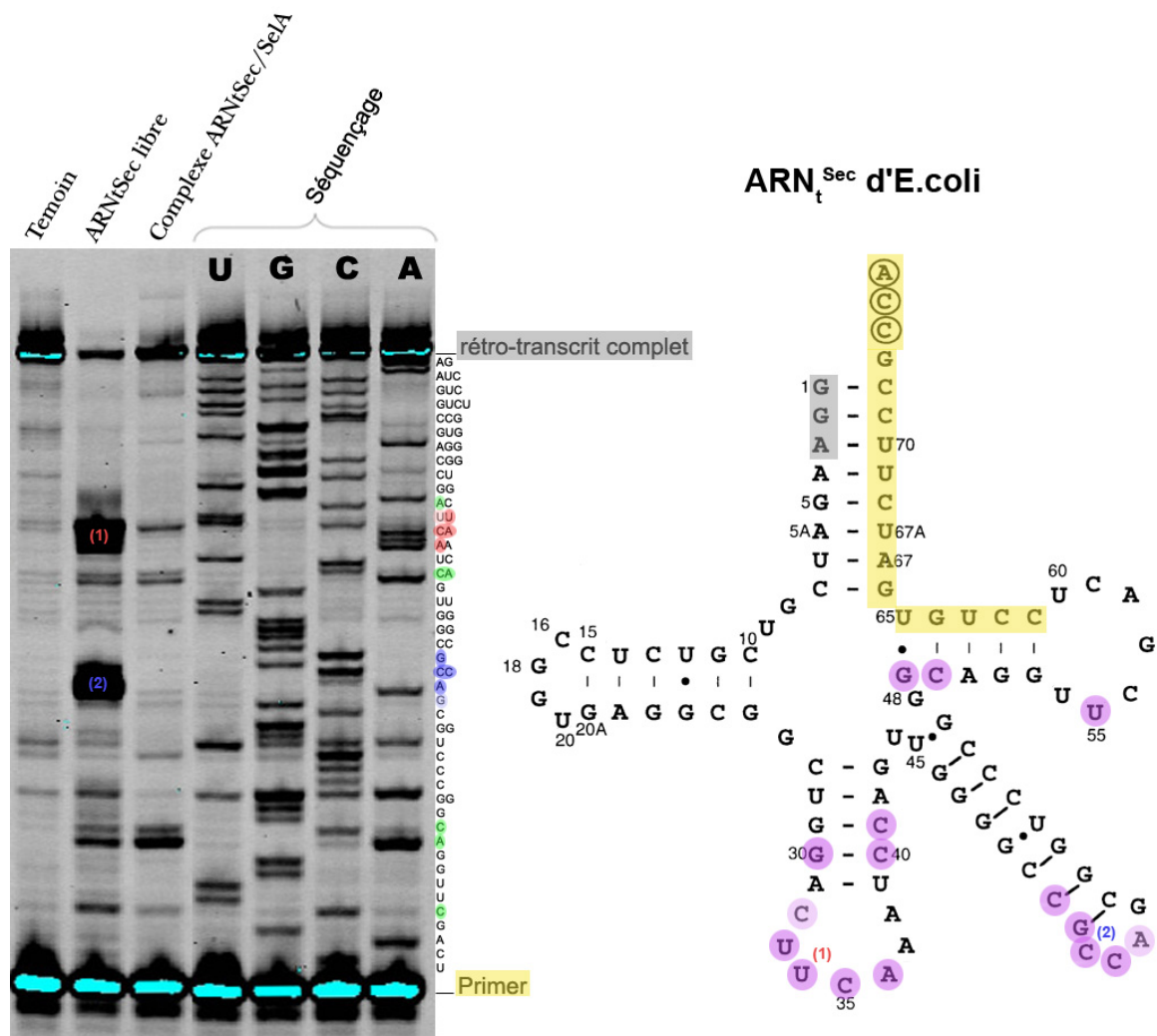


Figure 4.6 – Résultats de la méthode SHAPE sur l'ARN<sub>t</sub><sup>Sec</sup> libre et complexé à SslA.

Pour rappel, les nucléotides qui réagissent fortement lors de la réaction d'empreintes sur l'ARN<sub>t</sub><sup>ASP</sup> de *Saccharomyces cerevisiae*<sup>[II.39]</sup> sont : G18, U19, G30, U33, G34, U35, C36, G41, U55, G73 et A76, l'ARN<sub>t</sub><sup>ASP</sup> ne possédant pas de tige boucle variable (la numérotation classique utilisée permet une comparaison directe entre l'ARN<sub>t</sub><sup>ASP</sup> et l'ARN<sub>t</sub><sup>Sec</sup>).

L'ARN<sub>t</sub><sup>Sec</sup> présente une réponse à l'empreinte chimique au NMIA en grande partie conforme à celle attendue. Les déviations par rapport au résultat obtenu pour l'ARN<sub>t</sub><sup>ASP</sup> sont les suivantes :

- La boucle D de l'ARN<sub>t</sub><sup>Sec</sup> n'est pas réactive vis-à-vis du NMIA. Cela peut être expliqué par sa réduction à 4 nucléotides sur l'ARN<sub>t</sub><sup>Sec</sup>, contre 6 à 8 pour les ARN<sub>t</sub> classiques, ce qui résulte en des contraintes plus fortes sur ces nucléotides. Des contraintes dues aux interactions tertiaires formées entre la boucle D et la boucle T dans l'organisation tridimensionnelle de l'ARN<sub>t</sub> peuvent encore augmenter la rigidité de cette boucle.

- La réactivité des nucléotides G49 et C50 n'est pas explicable avec les informations dont nous disposons, ces deux nucléotides faisant partie d'un appariement double brin. Cette réactivité est à comparer à celle des nucléotides G30, C40 et C41, également impliqués dans un appariement, et qui possèdent, eux, des équivalents chez l' $\text{ARN}_t^{\text{Asp}}$  dans les nucléotides G30 et G41. La réactivité de ces derniers n'est d'ailleurs pas expliquée par les auteurs de l'analyse du comportement de l' $\text{ARN}_t^{\text{Asp}}$  face au réactif d'empreinte.<sup>[II.39]</sup>

Ce phénomène peut avoir une explication dans l'existence, dans certain cas, de nucléotides contraints dans une position où le phosphodiester en 3' est distant de l'hydroxyle 2'-OH, ce qui favorise la réaction de cette fonction avec le NMIA. L'exemple extrême d'une contrainte pouvant favoriser cette réaction est le cas de l'AMP cyclique, pour lequel le phosphodiester relie les extrémités 3' et 5' du nucléotide.<sup>[II.39]</sup>

- La boucle anticodon de l' $\text{ARN}_t^{\text{Sec}}$ , notée (1), est réactive sur plus de nucléotides que celle de l' $\text{ARN}_t^{\text{Asp}}$ , ce qui peut indiquer une plus grande flexibilité de celle-ci.

- La boucle de la tige-boucle variable de l' $\text{ARN}_t^{\text{Sec}}$ , notée (2), réagit également avec le réactif NMIA, ce qui rend compte de sa flexibilité et de l'absence d'interaction tertiaire avec le reste de l' $\text{ARN}_t^{\text{Sec}}$ .

Le résultat de l'expérience d'empreintes chimiques par SHAPE sur l' $\text{ARN}_t^{\text{Sec}}$  présent sous forme de complexe avec SelA montre une diminution, voire une disparition, de la réactivité de certains de ses nucléotides :

- La réactivité des bases C40, C41, G49, C50 et U55 reste quasiment inchangée.
- Par contre, mis à part le nucléotide U36, les boucles anticodon (1) et variable (2) ne réagissent plus à l'attaque chimique de l'anhydride NMIA.

### 4.3.3 Implications des résultats

Les déterminants de l' $\text{ARN}_t^{\text{Sec}}$  vis-à-vis des différents acteurs de la biosynthèse et de l'incorporation de la sélénocystéine ont été intensément étudiés (voir le § 2.1.1, page 82), et les résultats de ces études sont résumés sur l'alignement de la figure 4.7, page ci-contre.

Or les résultats obtenus par l'expérience d'empreintes chimiques SHAPE sur l' $\text{ARN}_t^{\text{Sec}}$  d'*E. coli* impliqué dans un complexe avec SelA — il s'agit ici de mtSelA — montrent que les boucles anticodon et variable, flexibles dans l' $\text{ARN}_t^{\text{Sec}}$  libre, ne le sont plus dans le complexe avec SelA.

Ces boucles semblent impliquées dans l'interaction de l' $\text{ARN}_t^{\text{Sec}}$  avec SelA. Pourtant le seul élément d'identité de l' $\text{ARN}_t^{\text{Sec}}$  à priori reconnu par SelA est sa tige acceptrice de 8 paires de bases.<sup>[II.37]</sup> Selon les études effectuées sur les discriminants de l' $\text{ARN}_t^{\text{Sec}}$  vis-à-vis de SelA, la réaction n'est pas remise en cause par le remplacement de l'anticodon UCA par un autre anticodon sur l' $\text{ARN}_t^{\text{Sec}}$ , ni par le remplacement de la tige-boucle variable par celle de l' $\text{ARN}_t^{\text{Ser}}$  à anticodon TGA,<sup>[II.37]</sup> pourtant moins longue (voir la figure 4.7, page suivante).

Il est donc possible que cette interaction entre SelA et l' $\text{ARN}_t^{\text{Sec}}$  ne soit pas spécifique de la séquence de l'ARN mais qu'il s'agisse d'une reconnaissance du squelette ribose-

DZ1660 E. Coli Sec (se/C)	DZ1430 M. Thermoacetica Sec (se/C)	DS1661 E. Coli Ser	DS1660 E. Coli Ser	DS1664 E. Coli Ser	DS1663 E. Coli Ser	
1 G	1 G	1 G	1 G	1 G	1 G	
2 G	2 G	2 G	2 G	2 G	2 G	
3 A	3 G	3 A	3 A	3 T	3 T	
4 A	4 G *	4 A	4 G	4 G	4 G	
5 G	5 G	5 G	5 A	5 A	5 A	
5a A	5a T					Tige acceptrice de 8pb
6 T	6 G	6 T *	6 G	6 G	6 G	
7 C	7 G	7 G	7 A	7 G	7 G	Discriminant pour EF-Tu
8 G	8 A	8 T	8 T	8 T	8 T	Changement pour U sans conséquence
9 T	9 T	9 G	9 G	9 G	9 G	Ecart par rapport aux ARNtSer d'E.Coli
10 C	10 C	10 G	10 C	10 T	10 G	
11 G	11 G *	11 C	11 C	11 C	11 C	Changement pour une paire C-G sans conséquence
12 T	12 G *	12 C	12 G	12 C	12 C	
13 C	13 G	13 G	13 G	13 G	13 G	Tige D apparée sur 6pb
14 T	14 C	14 A	14 A	14 A	14 A	Ecart par rapport au consensus
15 C	15 C	15 G	15 G	15 G	15 G	Ecart par rapport au consensus
16 C	16 T	16 C	16 C	16 T	16 A	
18 G	18 G	18 G	18 G	18 G	18 G	Position consensus chez les ARNt
19 G	19 G	19 G	19 G	19 G	19 G	Position consensus chez les ARNt
20 T	20 T	20 T	20 C	20 C	20 C	
20:B G	20:B G	20:A T	20:A T	20:A T	20:A T	Tige D apparée sur 6pb
21 A	21 G	21 A	21 A	21 A	21 A	Position consensus chez les ARNt
22 G	22 C	22 A	22 A	22 A	22 A	Tige D apparée sur 6pb
23 G	23 T *	23 G	23 C	23 G	23 G	
24 C	24 T *	24 G	24 G	24 G	24 G	Changement pour une paire C-G sans conséquence
25 G	25 G	25 C	25 G	25 A	25 C	
26 G	26 C	26 A	26 A	26 G	26 G	
27 C	27 C	27 C	27 C	27 C	27 C	
28 T	28 T	28 C	28 C	28 A	28 T	
29 G	29 G	29 G	29 G	29 C	29 C	
30 G	30 G	30 G	30 G	30 G *	30 C	
31 A	31 A	31 T	31 T	31 C	31 C	Ecart par rapport aux ARNtSer d'E.Coli
32 C	32 C	32 C	32 C	32 C	32 C	
33 T	33 T	33 T	33 T	33 T	33 T	
34 T	34 T	34 T	34 C	34 G	34 G	Changement pour G -> 30% d'incorporation
35 C	35 C	35 G	35 G	35 G	35 C	Changement de l'anticodon ( pour UGA) sans conséquence si le condon est changé conjointement
36 A	36 A	36 A	36 A	36 A	36 T	
37 A	37 A	37 A	37 A	37 A	37 A	
38 A	38 A	38 A	38 A	38 A	38 A	Ecart par rapport aux ARNtSer d'E.Coli
39 T	39 T	39 A	39 A	39 G	39 G	
40 C	40 C	40 C	40 C	40 T *	40 G	
41 C	41 C	41 C	41 C	41 G	41 G	
42 A	42 A	42 G	42 G	42 T	42 A	
43 G	43 G	43 G	43 G	43 G	43 G	
44 T	44 T	44 C	44 A	44 T	44 T	
e11 T *	e11 G = (45)	e11 G *	e11 G = (45)	e11 A = (45)	e11 A = (45)	
e12 G =	e12 C = (46)	e12 A =	e12 T = (46)	e12 T =	e12 T = (46)	
e13 G =	e13 G = (47)	e13 C =	e13 A = (47)	e13 A =	e13 G = (47)	
e14 G *	e14 C =	e14 C =	e14 G =	e14 C =	e14 C =	
e15 G *	e15 G =	e15 C =	e15 G *	e15 G =	e15 G =	
e16 C =	e16 C =		e16 G =		e16 G =	Ecart par rapport aux ARNtSer d'E.Coli
e17 C =	e17 G =				e17 T *	Tige variable particulièrement longue
e18 G =	e18 G =					Tige variable particulièrement longue
e19 G =						Tige variable particulièrement longue
e1 C	e1 C	e1 G	e1 G	e1 G	e1 C	
e2 C	e2 U	e2 A	e2 C	e2 C	e2 A	
e3 A	e3 A	e3 A	e3 A	e3 A	e3 A	
e4 G	e4 A	e4 A	e4 A	e4 A	e4 A	
e5 C					e5 A	
e28 C =	e28 C =					Tige variable particulièrement longue
e27 G =	e27 C =					Tige variable particulièrement longue
e26 G =	e26 G =					Tige variable particulièrement longue
e25 T *	e25 G =	e25 G =	e25 T *	e25 C =	e26 C =	Ecart par rapport aux ARNtSer d'E.Coli
e24 C =	e24 G =	e24 G =	e24 C =	e24 G =	e25 T =	
e23 C =	e23 C =	e23 G =	e23 T =	e23 T =	e24 C =	
e22 T =	e22 G =	e22 T =	e22 A =	e22 A =	e23 C =	
e21 G *	e21 C =	e21 T *	e21 C =	e21 T =	e22 A =	
48 G	48 G	48 C	48 C	48 C	e21 T =	
49 G *	49 G =	49 C	49 G	49 G		Ecart par rapport au consensus
50 C	50 T	50 A	50 G	50 G		Discriminant pour EF-Tu
51 A	51 G	51 G	51 G	51 G		Discriminant pour EF-Tu
52 G	52 G	52 A	52 G	52 G		Position consensus chez les ARNt
53 G	53 G	53 G	53 G	53 G		Position consensus chez les ARNt
54 T	54 T	54 T	54 T	54 T		Position consensus chez les ARNt
55 T	55 T	55 T	55 T	55 T		Position consensus chez les ARNt
56 C	56 C	56 C	56 C	56 C		Position consensus chez les ARNt
57 G	57 G	57 G	57 A	57 G		Position consensus chez les ARNt
58 A	58 A	58 A	58 A	58 A		Position consensus chez les ARNt
59 C	59 T	59 A	59 A	59 A		Ecart par rapport aux ARNtSer d'E.Coli
60 T	60 T	60 T	60 T	60 T		Position consensus chez les ARNt
61 C	61 C	61 C	61 C	61 C		Position consensus chez les ARNt
62 C	62 C	62 T	62 C	62 C		Position consensus chez les ARNt
63 T	63 C	63 C	63 C	63 C		
64 G	64 A	64 T	64 C	64 C		Discriminant pour EF-Tu
65 T	65 C	65 G	65 C	65 C		Discriminant pour EF-Tu
66 G	66 A	66 C	66 T	66 C		Discriminant pour EF-Tu
67 A	67 C	67 G *	67 C	67 C		
67a T	67a A					Tige acceptrice de 8pb
68 C	68 C	68 C	68 T	68 T	68 T	
69 T	69 T *	69 T	69 C	69 C	69 C	
70 T	70 C	70 T	70 T	70 A	70 A	
71 C	71 C	71 C	71 C	71 C	71 C	
72 C	72 C	72 C	72 C	72 C	72 C	
73 G	73 G	73 G	73 G	73 G	73 G	
74 C	74 C	74 C	74 C	74 C	74 C	
75 C	75 C	75 C	75 C	75 C	75 C	Position consensus chez les ARNt
76 A	76 A	76 A	76 A	76 A	76 A	Position consensus chez les ARNt

Figure 4.7 – Alignement de séquences des gènes <sup>187</sup>ARN<sub>t</sub><sup>Sec</sup> d'*E. coli* et de *M. thermoacetica*, et des ARN<sub>t</sub><sup>Ser</sup> isoaccepteurs d'*E. coli* ; et résumé des spécificités et discriminants des ARN<sub>t</sub><sup>Sec</sup> présentés.

phosphate de cet  $ARN_t$ . Cela rejoint les observations effectuées lors de la purification des protéines recombinantes Sela : la surface de Sela doit posséder des sites chargés positivement, interagissant avec les phosphates du squelette de l' $ARN_t^{Sec}$  de façon spécifique, ou d'autres fragments de matériel nucléique de façon non-spécifique (voir § 3.3.3.4, page 150). Il est donc peu probable que l'interaction de Sela avec l' $ARN_t^{Sec}$  ne se fasse qu'au niveau de la tige acceptrice, comme cela est le cas dans le modèle d'interaction entre l' $ARN_t^{Cys}$  et SepCysS chez certaines archaées<sup>[II.43]</sup> (voir la figure 1.23, page 59, ou 2.10, page 100).

Nous avons donc ici la preuve de l'interaction spécifique de Sela avec le squelette ribose-phosphate des boucles anticodon et variable de l' $ARN_t^{Sec}$ , grâce à la technique SHAPE. La portée de ces résultats sera discutée plus avant au § 5.1.2, page 208.

Les résultats obtenus par cette technique peuvent encore être améliorés, la fluorescence des brins rétrotranscrits donnant accès à leur quantification, ce qui permet d'accéder à la réactivité locale de l'ARN et donc à sa flexibilité. Les présents résultats sont surtout qualitatifs.

## 4.4 Cinétiques de la réaction de complexation

### 4.4.1 Rappel de la problématique

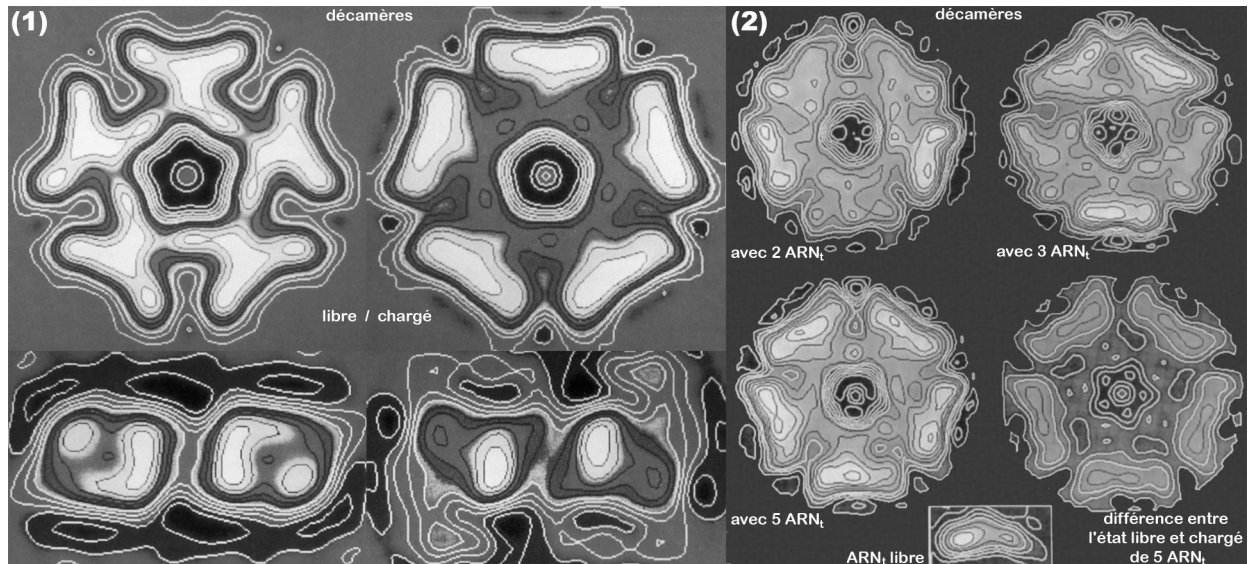
Des données sont actuellement disponibles sur le comportement enzymatique de Sela, en particulier sur la stoechiométrie de la complexation de Sela avec l' $ARN_t^{Sec}$  et sur la répartition de ces  $ARN_t^{Sec}$  sur un décimère, informations obtenues par analyse chromatographique, ou grâce à des clichés de cryomicroscopie électronique (voir 2.3.4, page 94) :

- Malgré la présence sur l'enzyme de 10 sites actifs contenant un cofacteur pyridoxal-5'-phosphate PLP, seuls 5 sites peuvent être conjointement occupés par l'aminoacyl porté par l' $ARN_t^{Sec}$ .
- Les 5 sites occupés sont portés par 5 différents dimères, en forme de T, du décimère : aucun dimère ne peut interagir avec 2 séryl- $ARN_t^{Sec}$ .
- L'état de charge complet du décimère, soit 5 sites occupés, n'est atteint que pour un large excès en séryl- $ARN_t^{Sec}$ , c-à-d. 6 fois les proportions stoechiométriques (mise en présence de 3 séryl- $ARN_t^{Sec}$  pour 1 monomère de Sela).
- Les états intermédiaires de charge du complexe, qui peuvent être observés par cryomicroscopie électronique, sont caractérisés par des remplissages des dimères 1 & 3 pour un décimère doublement chargé, et 1, 3 & 5 pour un décimère triplement chargé — la position de départ et le sens de la numérotation des sous-unités sont permutable, en raison des équivalences induites par les symétries de la protéine (voir le schéma 4.8, page suivante).
- Même si une piste existe à ce sujet (voir § 2.3.4.1, page 96), les caractéristiques de l'organisation de la répartition des  $ARN_t^{Sec}$  sur des dimères adjacents, et avec les autres dimères du complexe (organisation par alternance dessus/dessous, occupation d'une même face, organisation erratique), ne sont pas cernées actuellement.

Ces données ont permis aux équipes travaillant sur le sujet, dont celle de A. Böck,<sup>[II.9]</sup> d'émettre des hypothèses quant au fonctionnement de Sela. Selon cette dernière, il existe une coopérativité négative entre les deux sites présents sur une sous-unité dimérique, ce qui peut en effet être admis au regard des éléments précédents.

L'équipe conclut également qu'il n'y a pas de forte coopérativité entre les différents dimères d'un décimère, ni négative, ni positive — ce qui aboutirait respectivement à l'existence majoritaire de décimères chargés d'un seul  $\text{ARN}_t^{\text{Sec}}$ , ou de décimères chargés de cinq  $\text{ARN}_t^{\text{Sec}}$ . Existe-t-il alors tout de même une certaine coopérativité, faible, qui détermine une organisation du complexe, ou la répartition des  $\text{ARN}_t^{\text{Sec}}$  sur le décimère de Sela se fait-elle de façon aléatoire ?

L'observation d'états intermédiaires de complexation identifiables en cryomicroscopie électronique tend à réfuter la possibilité d'une organisation erratique du complexe.



**Figure 4.8** – Projections de clichés de cryo-microscopie électronique du complexe.<sup>[II.9]</sup>

(1) Comparaison entre l'enzyme libre (à gauche) et l'enzyme chargée de 5 seryl- $\text{ARN}_t^{\text{Sec}}$  (à droite). (2) Différents états de charge du complexe, et différence entre les états 0 et 5. En encart, cliché de l' $\text{ARN}_t^{\text{Sec}}$  seul, échelle respectée (75Å/30Å).

Ces états intermédiaires, à 2 ou à 3  $\text{ARN}_t^{\text{Sec}}$  sur un décimère, pourraient s'organiser de deux façons différentes :

- soit les  $\text{ARN}_t^{\text{Sec}}$  sont présents sur des dimères adjacents, ce qui n'est jamais observé,
- soit ils se répartissent pour mettre le plus de distance possible entre eux, ce qui est l'organisation observée.

(La position de l' $\text{ARN}_t^{\text{Sec}}$  considérée ne tient pas compte de la face d'interaction sur le dimère.)

Ce comportement peut être dépendant des propriétés de Sela, ou simplement découler de la gêne stérique ou la répulsion électrostatique entre un nouvel  $\text{ARN}_t^{\text{Sec}}$  interagissant avec la protéine et le ou les  $\text{ARN}_t^{\text{Sec}}$  déjà en place sur le décimère.

Dans les deux cas, une différence de paramètres cinétiques et thermodynamiques doit être mesurable entre la réaction d'interaction du premier séryl- $ARN_t^{Sec}$  sur un décacèle protéique libre et les réactions de liaison de chaque séryl- $ARN_t^{Sec}$  supplémentaire.

La connaissance de ces différents paramètres cinétiques permettrait de modéliser le comportement enzymatique de SelA, ce qui serait un "plus" pour notre étude.

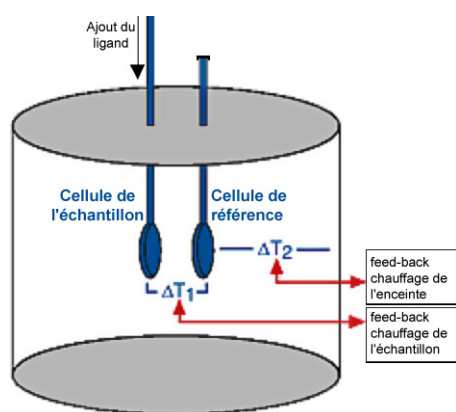
Il est également intéressant de mesurer les données cinétiques et thermodynamiques relatives à l'interaction de l' $ARN_t^{Sec}$  non-aminoacylé sur SelA, pour déterminer la force de l'interaction entre la protéine et le squelette ribose-phosphate de l' $ARN_t^{Sec}$ , et ainsi dévoiler un peu plus le fonctionnement de la formation du complexe SelA /  $ARN_t^{Sec}$ .

#### 4.4.2 Techniques de mesure

La laboratoire a acquis en 2007 un appareil de calorimétrie par titration isotherme. Je me suis donc intéressé à cette technique de mesure des paramètres cinétiques et thermodynamiques d'une réaction enzymatique.

La calorimétrie par titration isotherme (ITC) permet, en effet, de déterminer les paramètres caractérisant une interaction d'une macromolécule avec un ligand, notamment la constante d'association et la stoechiométrie de l'interaction, ainsi que les variations d'enthalpie ( $\Delta H$ ), d'entropie ( $\Delta S$ ) et de chaleur spécifique ( $\Delta C_p$ ) de l'interaction — l'obtention de cette dernière donnée demandant la réalisation de plusieurs mesures à des températures différentes.

L'appareil que j'ai utilisé pour l'étude des réactions entre SelA et l' $ARN_t^{Sec}$  est un microcalorimètre à compensation thermique (le modèle VP-ITC de Microcal Inc). Ce type d'appareil possède deux cellules constituées d'un alliage à haute conductivité thermique



**Figure 4.9** – Principe de l'appareil d'ITC - vue de l'enceinte thermostatée.

(Hastelloy®) : l'une d'elles contient la solution de référence, et l'autre contient le milieu réactionnel. Un premier dispositif thermoélectrique mesure la différence de température entre ces deux cellules et un deuxième dispositif mesure la différence de température entre les cellules et l'enceinte thermostatée de l'appareil maintenue à une température expérimentale donnée.

Si de la chaleur est produite ou absorbée dans la cellule réactionnelle, la différence de température entre cette dernière et la cellule de référence est immédiatement compensée électriquement. Cette puissance électrique, nécessaire au maintien de la température constante dans la cellule réactionnelle est enregistrée en continu. Elle correspond au signal expérimental brut et s'exprime en  $\mu\text{cal.s}^{-1}$  (voir figure 4.10, page suivante).

Dans une expérience typique, la macromolécule placée dans la cellule de mesure du microcalorimètre est progressivement saturée, à température imposée, par injections incrémentales de ligand délivrées à partir de la seringue motorisée de l'appareil, pilotée par ordinateur et équipée d'une hélice d'agitation. Chaque addition de ligand est accompagnée d'un dégagement thermique correspondant à l'interaction macromolécule / ligand.

L'ampleur de ce dégagement de chaleur est proportionnelle à la quantité de complexe formé, tant que la macromolécule titrée est en excès. Une fois que celle-ci arrive à saturation, le dégagement ou l'absorption de chaleur mesuré ne correspond plus qu'à la dilution de la solution concentrée de ligand dans le milieu réactionnel.

Si l'état d'équilibre de la réaction est rapidement atteint, le signal enregistré a l'aspect d'un pic, l'appareil corrigeant en temps réel son apport de chaleur au système pour conserver la même température dans les deux cellules. La quantité de chaleur totale résultant d'une injection correspond à l'aire du pic, c-à-d. l'intégrale par rapport au temps du signal expérimental. Le graphique montrant l'ensemble des quantités de chaleur décrivant une expérience de titration est appelé l'isotherme de liaison (en bas de la figure 4.10).

La connaissance des équations théoriques, décrivant la cinétique des interactions simples, a permis la programmation d'algorithmes de calcul, utilisés pour trouver les paramètres cinétiques et thermodynamiques correspondant le mieux aux mesures.

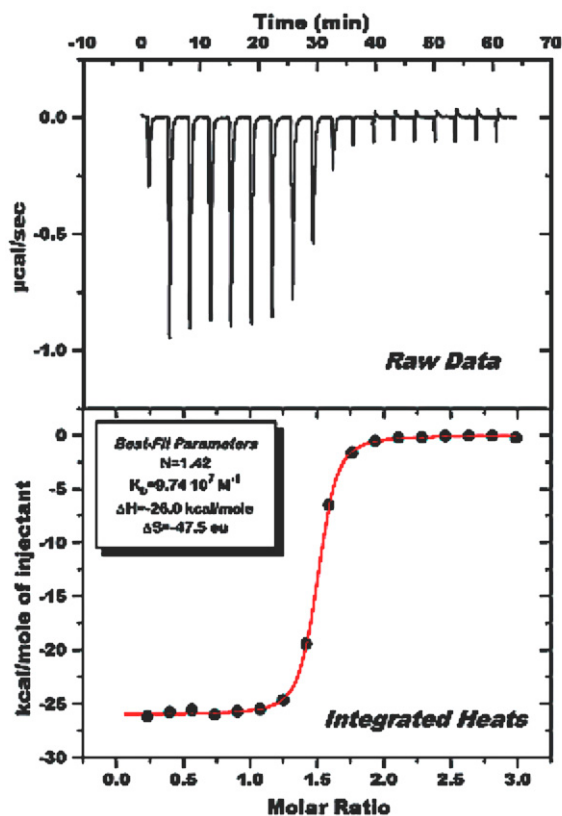


Figure 4.10 – Exemple de données mesurées et de modélisation, provenant de Microcal Inc.

Dans le cas où le site de fixation sur la macromolécule est unique, ou encore quand les sites multiples sont équivalents, l'isotherme de liaison obtenue est une sigmoïde, dont le point d'inflexion donne accès au coefficient stoechiométrique et dont la pente au point d'inflexion est proportionnelle à la constante d'association du complexe. La modélisation donne accès à ces données et à d'autres paramètres de la réaction avec une grande précision.

Dans les cas plus complexes, à plusieurs sites d'interaction non-équivalents, des modèles existent, mais leur pertinence est limitée aux cas strictement équivalents.

En général, la modélisation de l'isotherme de liaison permet d'obtenir simplement les valeurs des paramètres cinétiques et thermodynamiques désirées, grâce aux données



fournies au système, comme la concentration de départ de la macromolécule dans la cellule et celle du ligand dans la seringue.

La sélénocystéine synthase SelA est un décimère comportant 10 sites d'interaction équivalents, sur la protéine libre. Comme ce décimère ne peut accepter que 5 molécules d' $ARN_t^{Sec}$  au maximum, réparties sur chacun des dimères en forme de T de la molécule, une première approximation peut permettre de simplifier ce système d'étude : nous considérons que nous avons affaire à un décimère possédant 5 sites d'interactions équivalents, soit un par dimère en forme de T.

Le cas de SelA reste néanmoins délicat. Deux hypothèses sur le déroulement de la complexation de l' $ARN_t^{Sec}$  sont envisageables : soit il y a une levée de dégénérescence des sites au moment de la première liaison d'un  $ARN_t^{Sec}$  au décimère, dans quel cas les sites restants ne sont plus équivalents au premier site de fixation, soit la dégénérescence des sites est conservée et tous les sites peuvent être considérés comme équivalents.

Les données dont nous avons connaissance sur la complexation de l' $ARN_t^{Sec}$  par SelA, notamment la répartition systématique des ARN sur le décimère, rendent la première hypothèse plus vraisemblable.

Pour vérifier cela et tenter d'obtenir des informations cinétiques et thermodynamiques sur la formation du complexe, Nous avons suivi cette réaction par microcalorimétrie.

### 4.4.3 Mesures effectuées et résultats obtenus

J'ai entamé les études préliminaires au suivi de la formation du complexe par microcalorimétrie en novembre 2007. La difficulté du cas présenté et le temps important nécessaire à la mise au point de cette méthode de mesure ont retardé l'obtention de résultats. C'est donc une étude préliminaire qui est présentée dans ce manuscrit.

Je me suis intéressé dans un premier temps au cas de l'interaction entre SelA et l' $ARN_t^{Sec}$  non-acylé. L'étude de ce cas est en effet un peu plus simple à mettre en oeuvre, car l'utilisation d' $ARN_t^{Sec}$  non-acylé ne présente pas d'incompatibilité avec le pH nécessaire à la stabilité de SelA. Ce point est important, car c'est précisément ce qui pose problème lors des essais de suivi par microcalorimétrie de la formation du complexe covalent entre SelA et le séryl- $ARN_t^{Sec}$ , sur lequel je reviendrais plus tard.

Le point critique du protocole est le suivant : pour obtenir un suivi de réaction dans de bonnes conditions et accéder à des résultats précis, il est nécessaire de disposer de la macromolécule et du ligand en solution dans un même tampon, au même pH. En effet, en cas d'utilisation de tampons différents, l'énergie dégagée ou absorbée lors d'une injection de ligand est en partie due à la dilution du tampon présent dans la seringue dans le tampon présent dans la cellule, et aux possibles réactions acido-basiques, lorsque ces tampons ne sont pas au même pH.

Le but que je me suis fixé, était d'obtenir des résultats quantitatifs sur la formation du complexe. Pour cela, il me fallait disposer d'une isotherme de liaison complète et effectuer sa modélisation de façon adéquate.

Les conditions de la mesure devaient donc suivre un certain nombre d'obligations :

- Comme l'isotherme de liaison attendue est composée d'une à cinq sigmoïdes s'agencant sur plus d'un ratio stoechiométrique de partenaires, la réaction doit mettre en jeu 1,2 à 1,5 fois la quantité stoechiométrique de ligand pour que la modélisation soit possible.
- Les volumes de la cellule (1,4534 mL) et de la seringue (300  $\mu\text{L}$ ) du microcalorimètre déterminent le rapport de concentration des deux partenaires qu'il est nécessaire d'utiliser pour arriver au ratio molaire global précédent.
- Le nombre d'injections possibles est limité par le volume total de la seringue, et par le minimum d'énergie produite (ou absorbée) par injection qui doit favoriser à un rapport signal / bruit de fond suffisamment important pour la validité de la mesure.
- Dans notre cas, la concentration de SelA doit rester proche de 15  $\mu\text{M}$ , comme cela avait été déterminé lors du suivi de la formation du complexe par chromatographie analytique sur gel à exclusion stérique (voir § 3.4.1, page 154).

Les premières mesures valides effectuées grâce au microcalorimètre montrèrent les éléments suivants :

- Les pics relatifs à l'injection de ligand sont tout d'abord positifs, avant de devenir négatifs (voir le haut de la figure 4.11, page 195). Cela indique que la dilution de l' $\text{ARN}_t^{\text{Sec}}$  est endothermique, puis que l'interaction entre l' $\text{ARN}_t^{\text{Sec}}$  et SelA, qui est le deuxième phénomène observé, est exothermique.
- L'énergie dégagée par l'interaction entre SelA et l' $\text{ARN}_t^{\text{Sec}}$  est faible : une injection de 1,46 nmol d' $\text{ARN}_t^{\text{Sec}}$  (20  $\mu\text{L}$  à 73  $\mu\text{M}$ ) provoque la libération de seulement 7  $\mu\text{cal}$  apparentes, avec un maximum instantané de 0,1  $\mu\text{cal/s}$  (voir la figure 4.11, page 195).

Ces valeurs sont toutefois affectées par l'énergie absorbée par la dilution de l' $\text{ARN}_t^{\text{Sec}}$  dans le milieu réactionnel en début d'injection, qui est malheureusement incompressible (les deux partenaires sont ici en solution dans le même tampon).

Descendre en dessous de cette valeur d'énergie dégagée en diminuant le volume d'injection ou la concentration en  $\text{ARN}_t^{\text{Sec}}$ , ferait courir le risque d'avoir un niveau de bruit important au point d'influencer les valeurs d'intégration des pics obtenus, et ainsi l'isotherme de liaison mesurée.

La mesure des caractéristiques de la réaction de complexation entre SelA et l' $\text{ARN}_t^{\text{Sec}}$  par microcalorimétrie rencontre donc un problème technique majeur : le système d'équation à résoudre pour effectuer une mesure dans de bonnes conditions pose plus de contraintes qu'il n'y a de possibilités de variations des différents paramètres de la manipulation.

J'ai donc décidé de sacrifier la résolution de l'isotherme de liaison au profit du rapport signal mesuré / bruit de fond de la réaction : seulement 15 injections sont effectuées, mais cela permettra tout de même de visualiser la tendance globale de l'isotherme de liaison.

La meilleure mesure effectuée est présentée sur la figure 4.11, page suivante : elle combine l'absence d'artefact de mesure, avec un très faible niveau de bruit de fond et une ligne de base qui n'a nécessité qu'une faible correction.

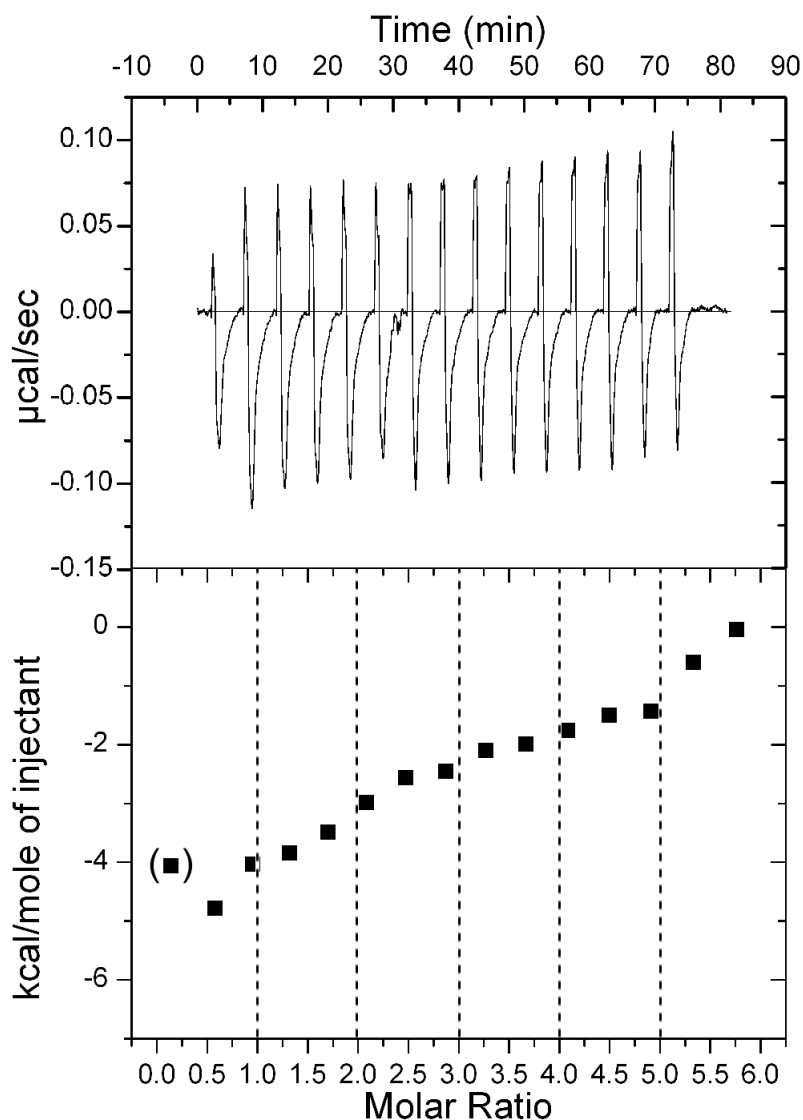
Les paramètres utilisés pour cette mesure sont les suivants :

- La température de l'isotherme de titration choisie est de 25°C.
- Les partenaires mis en présence sont mtSelA et l'ARN<sub>t</sub><sup>Sec</sup> de *Moorella thermoacetica*.
- Le tampon utilisé pour la mesure est le tampon de purification finale de SelA abaissé en pH, c-à-d. de l'HEPES 50mM à pH 7,3 contenant 100mM de KCl ; 80mM d'imidazole ; 20mM de MgCl<sub>2</sub> ; 1mM de DTT.
- La cellule de mesure contient 1,4534 mL de mtSelA à 24,3 μM, ce qui équivaut à 35,3 nmol de mtSelA ou 3,53 nmol de décamère mtSelA. La consigne de concentration maximale de SelA (15μM) a été ici un peu dépassée pour permettre d'avoir un bon ratio SelA / ARN<sub>t</sub><sup>Sec</sup>.
- La seringue contient 300 μL d'ARN<sub>t</sub><sup>Sec</sup> de *M. thermoacetica* à 73,1 μM. Les injections effectuées sont de 20μL, ce qui correspond à 1,46 nmol d'ARN<sub>t</sub><sup>Sec</sup> apportées au milieu par injection. Chaque injection correspond à un ratio molaire entre l'ARN<sub>t</sub><sup>Sec</sup> et le décamère de 0,41. C'est à ce niveau que le manque d'informations dégagées est le plus flagrant : la résolution de la courbe obtenue est très faible. La première injection n'est jamais prise en compte dans les calculs, car l'incertitude quant au volume réellement injecté est grande (la solution de ligand peut ne pas remplir toute la pointe de la seringue avant cette première injection). Le volume de la première injection a donc été réduit à 10 μL, ce qui a pour avantage de ne pas faire arriver la seringue en butée lors de la dernière injection.

L'isotherme obtenue ne permet pas d'obtenir immédiatement des valeurs numériques précises. Elle ne possède pas la forme d'une sigmoïde centrée sur un ratio molaire de 5 ARN<sub>t</sub><sup>Sec</sup> pour un décamère. On peut en conclure que les cinq interactions successives ne sont pas équivalentes (voir modélisation 4.4.4, page 197). Pour savoir si elles se différencient par leur énergie de réaction ou par la constante d'affinité du site d'interaction sur la protéine, une modélisation des résultats serait nécessaire. La vitesse d'association entre SelA et l'ARN<sub>t</sub><sup>Sec</sup> est assez faible : les injections de ligands nécessitent 6 minutes pour réagir entièrement avec la protéine présente dans le milieu, comme en témoigne la largeur des pics de mesure.

La méthode de calorimétrie par titration isotherme est une technique qui bénéficie d'une grande documentation scientifique,<sup>[II.44-II.47]</sup> plus d'un millier de publications (source : Microcal Inc.) comportent l'utilisation de cette technique pour la détermination des paramètres cinétiques et thermodynamiques d'une réaction d'interaction. Des exemples de modélisation d'interactions complexes à plusieurs sites actifs ont été publiés,<sup>[II.48-II.51]</sup> et portent sur des cas allant jusqu'à 4 sites actifs, identiques ou différenciés. Mais aucune de ces études ne porte sur l'interaction d'un ligand sur 5 sites actifs différenciés d'une macromolécule.

Je me suis donc intéressé à la modélisation des interactions entre SelA et l'ARN<sub>t</sub><sup>Sec</sup>, ce qui fera l'objet de la section suivante (§ 4.4.4, page 197).

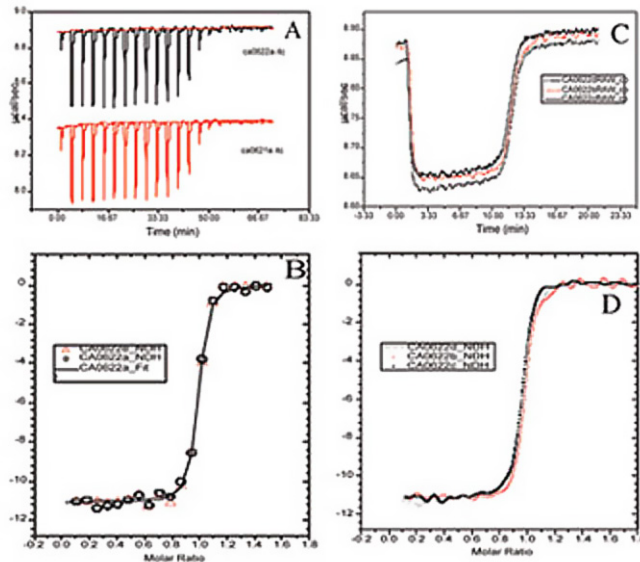


**Figure 4.11** – Suivi de la formation du complexe entre mtSelA et mtARN<sub>t</sub><sup>Sec</sup> par ITC. En haut, données brutes de mesure de microcalorimétrie, et en bas, isotherme de liaison obtenue (le ratio molaire correspond au rapport de la quantité d'ARN<sub>t</sub><sup>Sec</sup> introduite sur la quantité de **décamère** de SelA présente dans le milieu réactionnel).

Dans mes recherches de solutions à l'imprécision des mesures de l'expérience de microcalorimétrie — c-à-d., dans le cas présent, le manque de résolution de l'isotherme de liaison — je me suis intéressé à la méthode d'injection unique en continue du ligand ou SIM (Single Injection Method), par opposition à la méthode à injections multiples.

Je n'ai pas pu utiliser cette méthode avant la rédaction de ce manuscrit, mais la compréhension de sa philosophie m'a été utile lors des essais de modélisation de l'interaction étudiée.

Son principe est similaire à la méthode décrite précédemment, à la différence près que l'injection du ligand se déroule en une seule fois et en continu, sur toute la durée de la



**Figure 4.12** – Comparaison entre la méthode d'injections multiples et la méthode d'injection unique SIM. (A. et C. Les données brutes - B. et D. L'isotherme de liaison, respectivement) (Données provenant de Microcal Inc.)

mesure. L'isotherme de liaison est obtenue par intégration de la courbe de mesure par rapport à la quantité de ligand injectée par unité de temps, après la correction de la possible déviation de ligne de base de mesure.

La comparaison entre les deux méthodes, proposée par Microcal Inc. (voir figure 4.12), permet de vérifier que les isothermes de liaison obtenues pour les deux méthodes (B. pour les injections multiples et D. pour SIM) sont identiques. Cette image fournie dans la documentation de Microcal Inc. n'est pas de suffisamment bonne qualité pour le voir, mais les résultats quantitatifs sont décrits comme étant également identiques.

Par contre, cette méthode est plus sensible aux perturbations que la méthode à injection multiple, et demande sans doute plus de temps de mise au point. Il serait intéressant de vérifier si elle est applicable au cas de l'interaction de *SelA* avec  $ARN_t^{Sec}$ .

Dans le cas de l'interaction entre *SelA* et le séryl- $ARN_t^{Sec}$ , les divers essais effectués, avec la technique employée d'injections multiples, ne permirent pas d'obtenir d'isotherme de liaison exploitable.

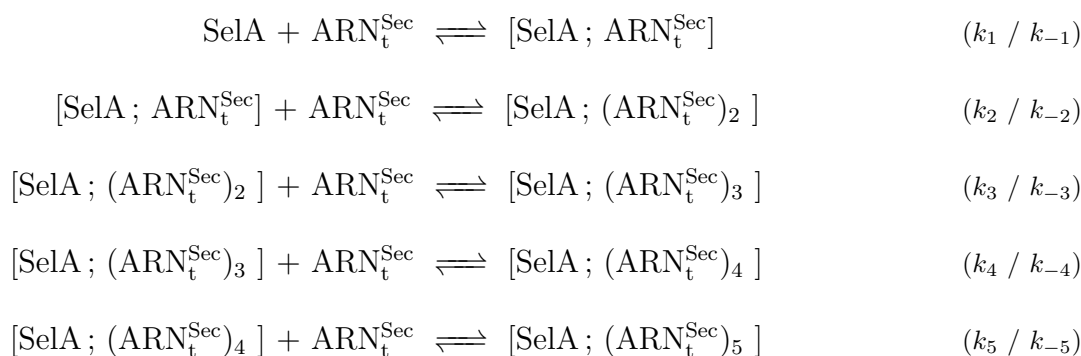
De plus, aucune solution n'a pu être trouvée pour le problème de compatibilité de tampon exposé précédemment. Cela peut contribuer au comportement erratique des mesures : soit la différence de tampon entre la solution du ligand et celle de la macromolécule provoque un dégagement de chaleur important par rapport au signal de l'interaction, soit la durée de l'expérience (3h30) implique un changement de la proportion d' $ARN_t^{Sec}$  déacylé entre le début et la fin de l'injection, à 25°C dans le tampon de réaction à pH 7,0.

Le projet de comparer les réactions d'interaction à différentes températures, pour les différents complexes possibles entre les versions de *SelA* recombinantes produites et les  $ARN_t^{Sec}$  d'*E. coli* et de *M. thermoacetica* n'a pu être entamé.

#### 4.4.4 Modélisation du comportement enzymatique de SelA

Le premier but de l'utilisation d'un logiciel de modélisation d'équilibre enzymatique et de réaction chimique fut de trouver un moyen d'interpréter les données obtenues par microcalorimétrie de titrage isotherme.

C'est donc sur le modèle simple suivant que je me suis basé, pour l'interaction entre SelA et l'ARN<sub>t</sub><sup>Sec</sup> non-acylé (SelA représente le décamère protéique, et les crochets représentent le complexe) :

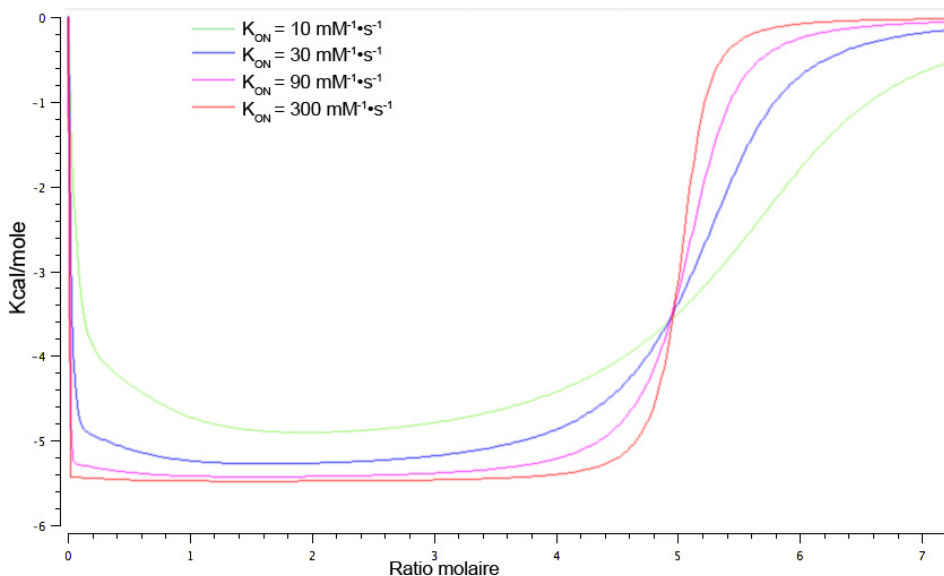


Pour modéliser ces équilibres combinés, j'ai utilisé l'outil logiciel puissant COPASI<sup>[II.52]</sup> développé grâce à une collaboration entre l'institut bioinformatique de Virginie, l'université d'Heidelberg et l'université de Manchester.

Ce logiciel est assez complexe à mettre en oeuvre, mais permet l'accès à la modélisation de la cinétique d'une réaction à partir de ses paramètres (calcul des stoechiométries massiques, caractérisation de l'état stationnaire, variation des populations des composés au cours de la réaction, vitesse de conversion instantanée selon l'avancement, ...).

Les données d'entrée nécessaires sont les équations du système (dans notre cas, présentées ci-dessus), les différentes constantes des réactions impliquées, ainsi que les volumes et concentrations du système à l'état initial.

Un premier essai de simulation est effectué avec les concentrations et volumes engagés lors de la mesure par calorimétrie, à 25°C. Un algorithme de contrôle de la quantité d'ARN<sub>t</sub><sup>Sec</sup> présente dans le milieu en fonction du temps permet de simuler une longue injection de ce ligand dans le milieu réactionnel (ce qui correspond à la méthode SIM).



**Figure 4.13** – Modèle de l'isotherme de liaison du système réactionnel précédent, avec chaque paramètre identique pour les cinq interactions :  $k_{-n} = 0,01 \text{ s}^{-1}$ ,  $k_n$  (soit  $k_{ON}$ ) est variable d'une courbe à l'autre et  $\Delta H = 5 \text{ kcal/mole}$  de ligand.

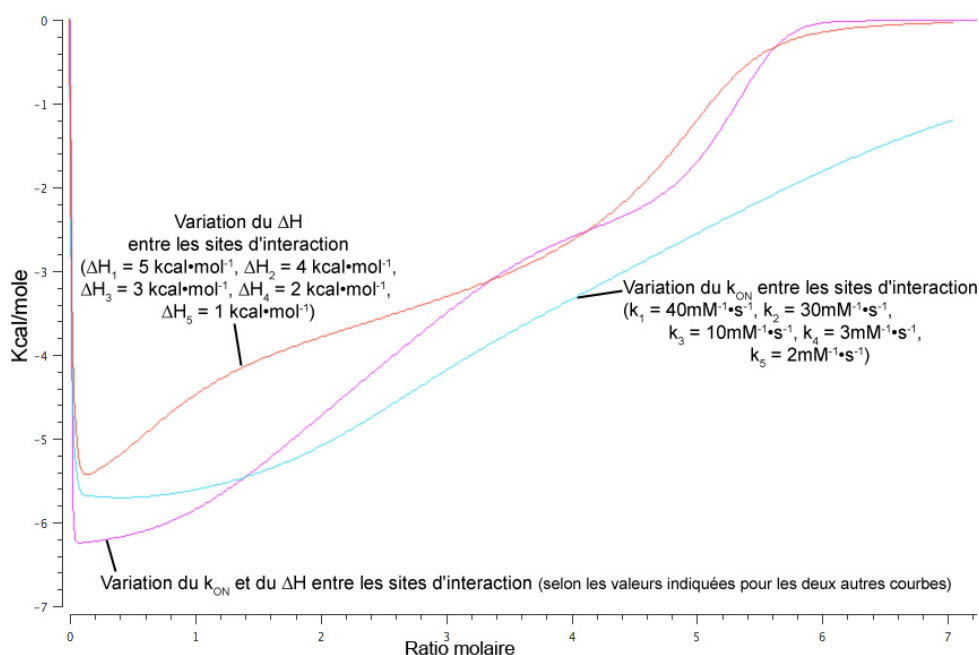
Les premières courbes obtenues grâce à la fonction de modélisation d'isothermes de liaison que j'ai programmée, pour différents paramétrages simples du système réactionnel, sont présentées sur la figure 4.13.

Ces premières courbes de modélisation permettent de se rendre compte que des sites d'interaction aux paramètres identiques se comportent de façon dégénérée : la situation est identique au cas d'une macromolécule ne possédant qu'un seul site, avec le point d'inflexion de la sigmoïde décalé au niveau du ratio molaire correspondant au nombre de sites dégénérés.

Cette première simulation confirme que les interactions successives entre Sela /  $ARN_t^{Sec}$  ne sont pas dégénérées : l'isotherme de titration obtenue par la méthode à injections multiples ne correspond pas à ce profil.

La question est de savoir si les sites de liaison successifs diffèrent par leur constante d'association ou simplement par la quantité d'énergie  $\Delta H$  dégagée lors de l'interaction.

Les modèles présentés sur la figure 4.14, page ci-contre montrent qu'une différence des sites d'interaction successifs, pour un des deux paramètres,  $\Delta G$  et  $k_{on}$ , ou pour les deux paramètres, modifie sensiblement l'isotherme de liaison obtenue. Il semblerait que ce soit sur les deux paramètres qu'il faille jouer pour obtenir un modèle correspondant à l'isotherme de liaison expérimentale de la figure 4.11, page 195



**Figure 4.14** – Modèle de l'isotherme de liaison du système réactionnel précédent, avec, pour chacune des cinq interactions,  $k_{-n} = 0,01 \text{ s}^{-1}$ ,  $k_n = 30 \text{ mM}^{-1}\cdot\text{s}^{-1}$  (c-à-d.  $k_{ON}$ ) et  $\Delta H = 5 \text{ kcal/mole}$  de ligand, sauf indication contraire.

Il s'agit donc de jouer sur 15 paramètres — les 5  $k_n$ , les 5  $k_{-n}$  et les 5  $\Delta H_n$  — et de comparer la courbe de simulation obtenue et la courbe expérimentale d'isotherme de liaison. Une légère modification d'un des paramètres et le calcul de la proximité du nouveau modèle avec la courbe expérimentale permet d'observer l'influence de chaque paramètre. Effectuée de façon itérative, cette méthode permet, si cela est possible, de faire converger des valeurs initiales vers un modèle solution.

Le logiciel de simulation d'équilibre COPASI n'est pas conçu pour ce genre de travail calculatoire, et étant donné le nombre important de paramètres, l'obtention d'un résultat convenable prit beaucoup de temps, une divergence du modèle pouvant vite s'installer.

Une amélioration de la modélisation pourra donner, dans un avenir proche, un moyen efficace de simulation et de calcul des paramètres cinétiques d'un système complexe, généralement non pris en compte dans les logiciels accompagnant les appareils d'ITC, à partir de ses isothermes de liaison.

Actuellement, le meilleur modèle de l'isotherme de liaison expérimentale obtenu est celui présenté à la figure 4.15, page 201. Ce modèle peut certainement encore être amélioré, mais donne tout de même un bon aperçu des paramètres aboutissant à la courbe expérimentale.

Les paramètres de ce modèle sont résumés dans le tableau 4.1, page 201.



On peut s'apercevoir que les valeurs des  $\Delta H_n$  et des  $k_{-n}$  sont raisonnables, alors que celles des  $k_n$  dépassent les possibilités :

Par exemple, pour l'injection de 20  $\mu\text{L}$  de ligand à 73  $\mu\text{M}$  dans 1,4534 mL d'une solution de décamère à 2,5  $\mu\text{M}$ , la réaction de formation du complexe (coefficient  $k_1$ ) s'effectuerait à la vitesse de 75  $\mu\text{mol}$  par seconde, et donc durerait 13  $\mu\text{s}$ !

Comment expliquer ces résultats ?

Le premier élément de réponse provient des données expérimentales obtenues : vu la faible précision de ces données expérimentales, certaines variations de l'isotherme de liaison sont sans doute plus artefactuelles que liées au comportement réel du complexe : le modèle tente de s'adapter à ces variations, d'où la génération de valeurs en contradiction avec le comportement réel du complexe.

L'autre possibilité, pour expliquer ces valeurs aberrantes, vient de la nature même du modèle : celui-ci tente de faire correspondre deux courbes par la variation de **15** paramètres. Il serait donc possible que la modélisation en question soit tombée dans un "puits de potentiel" et que d'autres solutions existent.

L'amélioration des connaissances sur la formation du complexe est possible, grâce à la multiplication des données expérimentales, et à l'amélioration de la méthode de modélisation. Ce cas de complexe à deux partenaires et 5 sites d'interactions étant particulièrement singulier, la littérature scientifique disponible ne nous apporte aucune aide pour la résolution de ce problème.

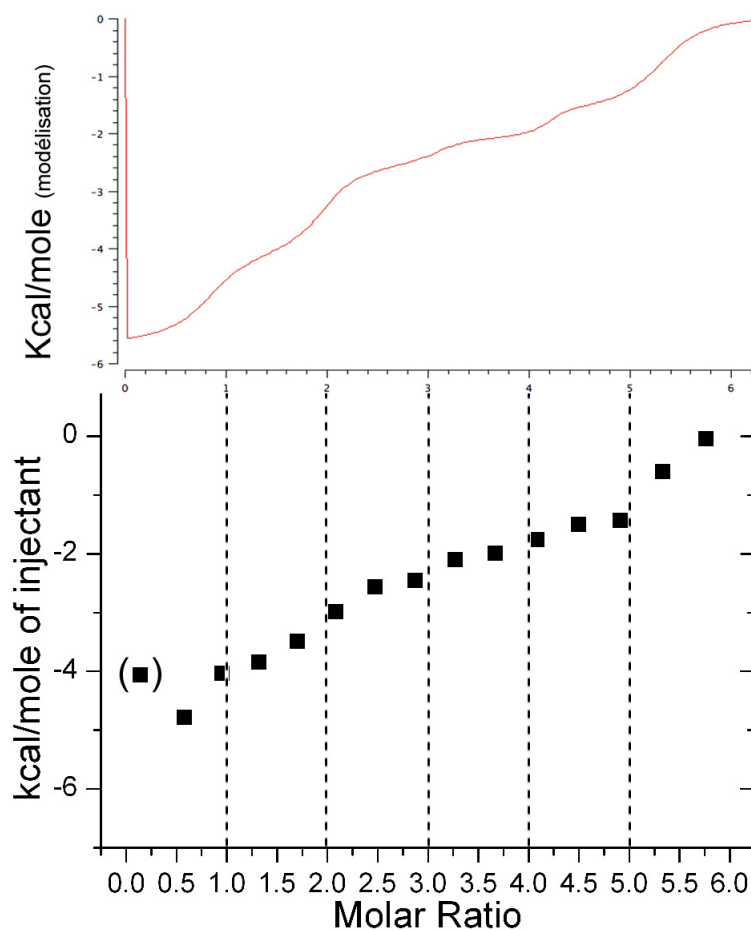
Bien que les données obtenues par modélisation ne puissent être considérées comme des indicateurs quantitatifs de l'interaction entre SelA et l' $ARN_t^{Sec}$ , la tendance globale de ces valeurs, à confirmer par l'amélioration du modèle, est tout de même notable : l'énergie dégagée par l'interaction et son coefficient d'association seraient des fonctions décroissantes du nombre d' $ARN_t^{Sec}$  déjà présents sur le décamère.

Le but premier de toute cette étude des interactions entre SelA et l' $ARN_t^{Sec}$ , à la fois expérimentale et calculatoire, était d'apporter des éléments sur la connaissance de la formation du complexe que nous avons tenté de cristalliser : celui de SelA et du seryl- $ARN_t^{Sec}$ .

En effet, le système de réactions combinées amenant au complexe utilisé en cristallogénèse descendant directement de celui existant entre SelA et l' $ARN_t^{Sec}$  non-acylé (présenté en début de ce chapitre, page 197). Il est cependant un peu plus complexe que le précédent, puisqu'il doit prendre en compte la formation de la liaison covalente entre l'amine de la sérine esterifiée à l' $ARN_t^{Sec}$ , avec le pyridoxal-5'-phosphate PLP présent dans le site actif de la macromolécule.

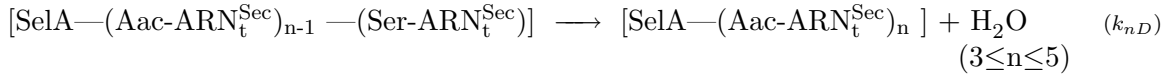
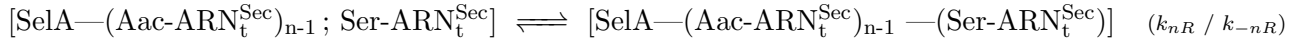
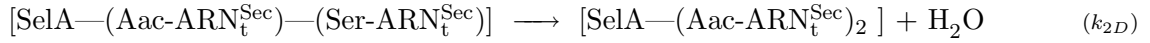
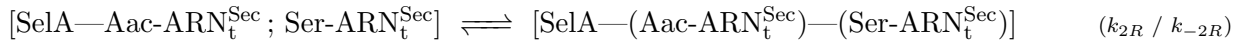
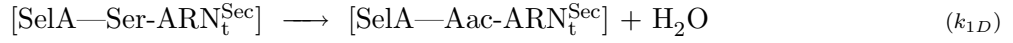
Paramètres	$\Delta H_n$	$k_n$	$k_{-n}$
Valeurs pour l'interaction n			
n = 1	5,55 kcal/mole	$3,0 \times 10^7 \text{ mM}^{-1} \cdot \text{s}^{-1}$	$4,0 \times 10^{-3} \text{ s}^{-1}$
n = 2	4,05 kcal/mole	$1,8 \times 10^6 \text{ mM}^{-1} \cdot \text{s}^{-1}$	$4,0 \times 10^{-3} \text{ s}^{-1}$
n = 3	2,55 kcal/mole	$9,0 \times 10^4 \text{ mM}^{-1} \cdot \text{s}^{-1}$	$4,5 \times 10^{-3} \text{ s}^{-1}$
n = 4	2,08 kcal/mole	$4,0 \times 10^3 \text{ mM}^{-1} \cdot \text{s}^{-1}$	$5,0 \times 10^{-3} \text{ s}^{-1}$
n = 5	1,62 kcal/mole	$1,0 \times 10^2 \text{ mM}^{-1} \cdot \text{s}^{-1}$	$6,0 \times 10^{-3} \text{ s}^{-1}$

**Tableau 4.1** – Paramètres du modèle présenté à la figure 4.15. (Les unités sont celles utilisées par le logiciel de simulation.)



**Figure 4.15** – Comparaison entre l'isotherme de liaison expérimentale (en bas) et son modèle (en haut) régi par les paramètres du tableau 4.1.

Voici le système réactionnel en question : (Ser = séryl et Aac = aminoacrylyl)



Nous décomposons donc l'interaction et la réaction d'un séryl-ARN<sub>t</sub><sup>Sec</sup> avec SelA en trois étapes :

- une première étape de “reconnaissance” du substrat, caractérisée par l'interaction de l'ARN<sub>t</sub><sup>Sec</sup> avec SelA,
- une seconde étape de liaison du séryl porté par l'ARN<sub>t</sub><sup>Sec</sup> au PLP,
- une troisième étape de déshydratation du séryl en aminoacrylyl.

Les réactions de dégradation du complexe covalent ne sont pas considérées dans le système ci-dessus, mais pourraient y être ajoutées, comme celle de formation de pyruvate (Pya) :



La possibilité qu'un nouveau séryl-ARN<sub>t</sub><sup>Sec</sup> se lie au décimère alors que ceux déjà en place ne sont pas encore déshydratés n'est pas prise en compte dans ce modèle simplifié, car cela résulterait en une complexification très importante du système.

Pour l'étude de ce système déjà très complexe, d'autres approximations ont été envisagées :

- Il est chimiquement vraisemblable que les  $k_{nD}$  et les  $\Delta H_n D$  associés à la réaction de déshydratation du séryl en aminoacrylyl soient très proches pour les différents  $n$  de 1 à 5, et puissent être considérés comme identiques.
- De même, les  $\Delta H_n R$  de la réaction de liaison entre le PLP et le séryl peuvent être considérés comme proches, alors que les  $k_{nR}$  et  $k_{-nR}$  sont laissés libres de rester dégénérés ou non.

Malgré les efforts effectués pour rendre ce système modèle proche des observations expérimentales (de façon qualitative), aucune modélisation n'arrive à rendre la nécessité de la présence d'un excès de 6 fois le ratio stoechiométrique en séryl-ARN<sub>t</sub><sup>Sec</sup> par rapport au décimère de Sela pour obtenir un complexe complètement chargé : tous les modèles montrent la formation du complexe complet au bout d'un certain temps, même avec peu d'excès de séryl-ARN<sub>t</sub><sup>Sec</sup>.

Cela était prévisible, vu que dans notre système, les équilibres réactionnels sont déplacés par l'irréversibilité de la réaction de déshydratation. Il semblerait néanmoins que cela ne soit pas le seul élément expliquant l'écart de ce modèle avec la réalité : le formation du complexe entre Sela et l'ARN<sub>t</sub><sup>Sec</sup> doit suivre un mécanisme dont certains points nous échappent encore actuellement.

Cette étude de la formation du complexe entre Sela et l'ARN<sub>t</sub><sup>Sec</sup> a permis d'apporter de nouveaux éléments sur le principe de fonctionnement de l'enzyme, comme la vérification d'une levée de dégénérescence des sites d'interaction du décimère lors de la fixation d'un ARN<sub>t</sub><sup>Sec</sup>, et donc l'existence d'une collaboration négative entre les sites de l'ARN<sub>t</sub><sup>Sec</sup>.

Pourtant, la complexité de ce phénomène, encore baigné d'inconnu, reste entière, et cela malgré les hypothèses et les modélisations effectuées.



# Chapitre 5

## Discussion sur la biosynthèse de la sélénocystéine chez les eubactéries

### Sommaire

---

<b>5.1</b>	<b>Portée des résultats expérimentaux obtenus . . . . .</b>	<b>206</b>
5.1.1	Au niveau structural... . . . . .	207
5.1.2	Analyse du fonctionnement enzymatique de SelA . . . . .	208
<b>5.2</b>	<b>Étude structurale théorique sur SelA . . . . .</b>	<b>210</b>
5.2.1	Calculs de prédictions des structures secondaires de SelA .	211
5.2.2	Comparaisons structurales avec les protéines homologues .	212
5.2.3	Vers un modèle structural de SelA ? . . . . .	219
<b>5.3</b>	<b>Perspectives . . . . .</b>	<b>220</b>

---

Au cours des précédents chapitres, nous avons parcouru les différentes évolutions méthodologiques et scientifiques qui ont marqué ce travail de thèse.

Même si un certain nombre de réponses aux problématiques que soulève la biosynthèse de la sélénocystéine n'ont pu être apportées, des avancées ont été réalisées dans la connaissance des interactions enzymatiques impliquant SelA et des conditions expérimentales nécessaires à l'obtention de cristaux contenant cette protéine.

De nombreuses pistes ont été ouvertes, qui mériteraient davantage de temps pour être totalement explorées et aboutir à des résultats plus concluants.

Ce chapitre propose un rappel des différents éléments qui ont pu être apportés par ce travail, et conclut par une discussion sur le mécanisme de biosynthèse de la sélénocystéine, enrichie d'une analyse récente de données bioinformatiques sur la structure de la sélénocystéine synthase.

## 5.1 Portée des résultats expérimentaux obtenus

Ce travail de doctorat a évolué suivant deux grands axes d'étude de l'étape dans la biosynthèse de la sélénocystéine que constitue le complexe entre SelA et l'ARN<sub>t</sub><sup>Sec</sup>.

Le premier axe d'étude fut orienté en direction de l'obtention de données structurales concernant ce complexe. Ce travail a été la base de tout mon apprentissage des méthodes et des problématiques de la biologie structurale, et en particulier des problématiques concernant les partenaires d'étude de mon sujet.

La mise au point des conditions de purification des constructions protéiques étudiées (§ 3.3.3.1, page 145), des conditions de formation et de purification du complexe (§ 3.4, page 153), puis des conditions de cristallogenèse de celui-ci (§ 3.5.2, page 160), ont permis d'aboutir à la formation de cristaux de ce complexe protéo-nucléique (figures 3.20 et 3.21, pages 169-171).

Ce cheminement, en but d'obtenir des connaissances structurales et fonctionnelles de SelA par diffraction aux rayons X d'un cristal de son complexe avec l'ARN<sub>t</sub><sup>Sec</sup>, permet également d'accéder à certaines données sur le comportement de cette protéine en solution. Nous y reviendrons au § 5.1.1, page suivante.

Le deuxième axe d'étude de ce travail découle de ces recherches d'information sur le comportement de ses macromolécules en solution. L'étude enzymatique de SelA a donc été amorcée par l'inventaire des particularités observées lors de la mise au point de sa purification. À la fin de ce travail de doctorat, cette étude enzymatique a été approfondie par des recherches plus poussées sur les interactions présentes dans le complexe SelA / ARN<sub>t</sub><sup>Sec</sup>, par empreintes chimiques (§ 4.3, page 182) et calorimétrie par titration isotherme (§ 4.4, page 188).

La portée des résultats obtenus par ces techniques sera abordée en détail au § 5.1.2, page 208.

Ces réflexions sur le comportement réactionnel de SelA ont permis de mettre au point un modèle d'étude de ce système, malheureusement encore éloigné de sa réalité. L'amélioration de cette modélisation demeure subordonnée à l'augmentation des connaissances sur ce phénomène et à la conduite de nouvelle expérimentation.

La rédaction de ce manuscrit, en particulier de la concaténation bibliographique que constitue la première partie, a également été propice à la réflexion sur les questions structurales et phylogéniques concernant SelA. Cette réflexion, étayée de données bioinformatiques obtenues très récemment, est présentée au § 5.2, page 210 et permet d'ouvrir de nouvelles pistes pour un développement futur de ce projet de recherche.

### 5.1.1 Au niveau structural...

Dans le but d'obtenir des cristaux du complexe entre SelA et l'ARN<sub>t</sub><sup>Sec</sup>, le processus de préparation des partenaires a subi d'importantes améliorations.

Malgré les efforts effectués dans ce sens, la production reproductible de cristaux des macromolécules d'étude est une étape qui n'a pu être atteinte. La structure de SelA n'est donc pas encore à notre portée.

Ce cheminement a tout de même permis d'aboutir aux observations suivantes :

- L'utilisation des montages protéiques ecSelA, mtSelA et mtSelA( $\Delta$ 1-31) comportant un tag à 6 histidines a le grand avantage de permettre une purification rapide de ces protéines, mais conduit à une stabilité en solution moins importante que celle des versions sauvages. Celle-ci nécessite en effet l'utilisation de tampons contenant un minimum de 80mM d'imidazole, et d'un pH basique. De telles précautions ne semblent pas nécessaires lors de l'utilisation des protéines sauvages.<sup>[II.9,II.10,II.14,II.31]</sup> Comme les problèmes d'insolubilité causés par le tag à 6 histidines sont connus, la coupure de ce tag à la thrombine a été tentée, mais les rendements de coupure étaient trop faibles pour pouvoir obtenir des protéines sans ce tag en quantité importante.
- Ces montages protéiques s'organisent bien sous forme de décimères et forment des complexes avec l'ARN<sub>t</sub><sup>Sec</sup>, comme le montre leur profil de migration en chromatographie sur gel à exclusion stérique : ni la présence du tag, ni la troncature de la partie N-terminale de la version mtSelA( $\Delta$ 1-31) ne perturbent le fonctionnement de SelA.
- Cette technique de chromatographie a également permis d'observer l'existence de trois populations de protéines recombinantes différentes. En particulier, l'existence de la protéine sous la forme de monomère, ayant sans doute un défaut d'organisation structurale, et d'agrégats pouvant inclure du matériel nucléaire, souligne la possibilité qu'un échantillon protéique pur ne soit pas homogène. La cristallogénèse entreprise sur SelA nécessite donc le passage par cette technique de purification.
- L'existence de protéines SelA sous forme d'agrégats contenant du matériel nucléaire permet de confirmer la présence sur cette protéine de sites de surfaces ayant une affinité pour du matériel nucléaire chargé négativement. Cela confirme la possibilité d'une interaction entre SelA et le squelette ribose-phosphate de l'ARN<sub>t</sub><sup>Sec</sup>.
- Lors de la formation du complexe *in vitro*, un problème de compatibilité des pH de stabilité des deux partenaires, SelA et l'ARN<sub>t</sub><sup>Sec</sup>, a pu être constaté. Ceci rend l'utilisation de ces partenaires délicate pour des études enzymatiques. Une solution à ce problème a été développée de manière à former ce complexe pour la cristallogénèse, ce qui a permis d'obtenir des cristaux de matériel protéique. Cette obtention de cristaux reste la preuve



que le processus mis au point pour la préparation des partenaires biologiques était une avancée nécessaire.

L'information manquante pour valider complètement ce processus de purification reste la diffraction de ces cristaux, donnée importante pour déterminer la viabilité d'une poursuite de ce projet en l'état.

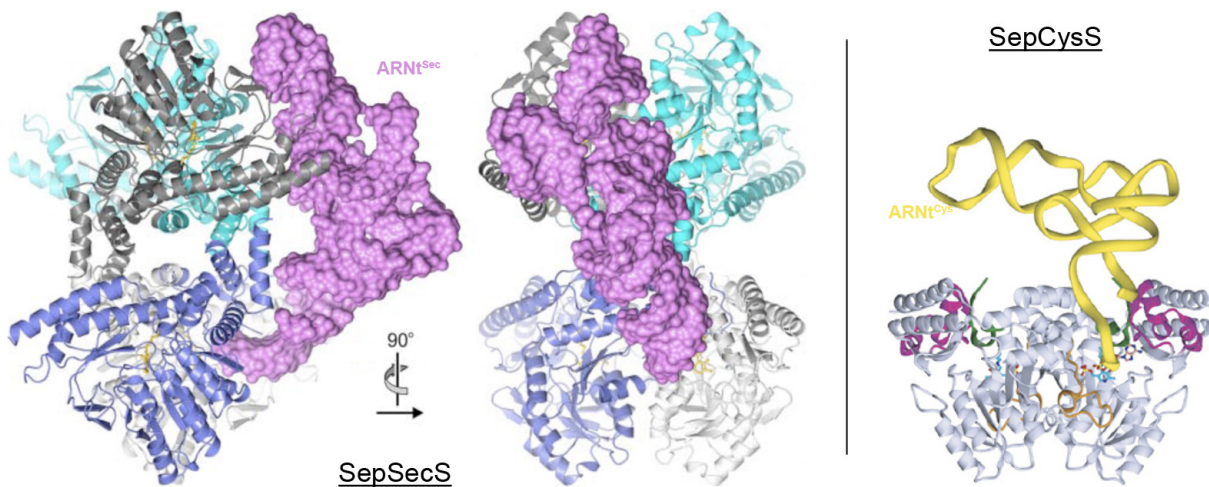
### 5.1.2 Analyse du fonctionnement enzymatique de SelA

Les connaissances acquises lors des deux premières années de ce travail de doctorat, à la fois sur les techniques employées en biologie structurale et sur les données disponibles concernant les partenaires à l'oeuvre dans le complexe d'étude, m'ont permis d'appréhender l'analyse du fonctionnement enzymatique de SelA dans des conditions optimales.

Grâce aux techniques d'empreintes chimiques par SHAPE et de microcalorimétrie par titrage isotherme, de nouvelles données sur le comportement enzymatique de SelA ont pu être dévoilées.

L'expérience par empreinte chimique SHAPE nous a permis de mettre en avant l'existence d'une interaction entre SelA et les boucles anticodon et variable de l' $\text{ARN}_t^{\text{Sec}}$ . Vu les données existantes sur les déterminants de l' $\text{ARN}_t^{\text{Sec}}$  nécessaires à sa réaction avec SelA,<sup>[II.30,II.53]</sup> cette interaction ne doit pas être spécifique à la séquence de l' $\text{ARN}_t^{\text{Sec}}$ , mais doit permettre une reconnaissance du squelette ribose-phosphate de l' $\text{ARN}_t^{\text{Sec}}$ .

Cette interaction nous permet d'effectuer un parallèle du fonctionnement du complexe SelA /  $\text{ARN}_t^{\text{Sec}}$  avec les modèles postulés pour les complexes SepSecS /  $\text{ARN}_t^{\text{Sec}}$ <sup>[II.3]</sup> et SepCysS /  $\text{ARN}_t^{\text{Cys}}$ <sup>[II.43]</sup> (voir § 2.3.5.2, page 100, et § 2.3.5.1, page 99, respectivement).



**Figure 5.1** – Modèles des complexes SepSecS /  $\text{ARN}_t^{\text{Sec}}$ <sup>[II.3]</sup> et SepCysS /  $\text{ARN}_t^{\text{Cys}}$ <sup>[II.43]</sup>.

Ces modèles présents sur la figure 5.1 montrent comment peut être envisagée, de deux façons très différentes, l'interaction entre l'aminoacyl-ARN<sub>t</sub> et l'enzyme PLP-dépendante qui modifie cet aminoacyl. (La différence de taille des ARN<sub>t</sub> provient de la différence de représentation utilisée, mais aussi du fait que l'ARN<sub>t</sub><sup>Sec</sup> de *Mus musculus*<sup>[II.3]</sup> compte 90 nucléotides, alors que l'ARN<sub>t</sub><sup>Cys</sup> d'*Archaeoglobus fulgidus*<sup>[II.43]</sup> en compte 75.)

Le complexe SelA / ARN<sub>t</sub><sup>Sec</sup> ne peut pas ressembler au modèle établi pour SepCysS, dans lequel aucune interaction n'existe entre les boucles anticodon de l'ARN<sub>t</sub><sup>Cys</sup> et la protéine. Dans le cas du complexe SepSecS, une telle interaction est postulée. Elle est même supposée être la raison de l'organisation oligomérique particulière de SepSecS en tétramère et de la distance existant entre les deux dimères de la protéine.<sup>[III.3]</sup>

SelA possède une structure oligomérique en décamère, unique chez les enzymes PLP-dépendantes. Il serait également envisageable que cette structure quaternaire soit en relation avec le mode d'interaction de la protéine avec l'ARN<sub>t</sub><sup>Sec</sup>. Cette hypothèse ne pourra être confirmée qu'avec la résolution de la structure de SelA ou de complexe.

Elle suppose néanmoins une organisation stricte des ARN<sub>t</sub><sup>Sec</sup> sur le décamère SelA : si l'interaction d'un ARN<sub>t</sub><sup>Sec</sup> se fait avec plusieurs dimères, un dimère est en liaison avec plusieurs molécules d'ARN<sub>t</sub><sup>Sec</sup> (au moins deux) dans le complexe totalement chargé. Une telle organisation associée à la symétrie d'ordre 5 observable sur les clichés de cryomicroscopie électronique du complexe (voir § 2.3.4.1, page 96) implique que tous les ARN<sub>t</sub><sup>Sec</sup> soient présents sur une seule face du décamère, et cela à cause de la chiralité de l'assemblage ARN<sub>t</sub><sup>Sec</sup> / dimère de SelA (principe de l'hélice à pales asymétriques).

Cette organisation poserait alors un problème d'encombrement important des dimères.

Il serait plus vraisemblable que les deux connexions entre chaque ARN<sub>t</sub><sup>Sec</sup> et SelA — celle au niveau du site actif contenant le PLP d'une part, et celle entre les boucles anticodon / variable de l'ARN<sub>t</sub><sup>Sec</sup> et les sites chargés positivement de SelA d'autre part — prennent place sur le même dimère.

Nous reviendrons sur ces hypothèses lors de l'analyse des prédictions de structure de SelA au § 5.2.3, page 219.

À côté de ces données sur les zones d'interaction de l'ARN<sub>t</sub><sup>Sec</sup> avec SelA, obtenues par empreinte chimique SHAPE, le fonctionnement de l'interaction entre SelA et l'ARN<sub>t</sub><sup>Sec</sup> a pu être en partie étudié par microcalorimétrie. L'incompatibilité des zones de pH de stabilité pour SelA et le séryl-ARN<sub>t</sub><sup>Sec</sup> a en effet rendu impossible l'obtention de résultats sur l'interaction de ce couple de partenaires. Des mesures ont par contre pu être effectuées pour l'interaction entre SelA et l'ARN<sub>t</sub><sup>Sec</sup> non acylé, révélant ainsi un aspect du mécanisme enzymatique de SelA.

L'isotherme de liaison obtenu pour cette interaction permet d'écarter la possibilité d'une dégénérescence entre les sites du complexe lors de l'interaction successive de plusieurs molécules d'ARN<sub>t</sub><sup>Sec</sup>.

Cette levée de dégénérescence va dans le sens d'une coopérativité négative entre les sites de SelA. D'autres mesures sont nécessaires pour confirmer la tendance observée

sur les récentes données obtenues : il semblerait que l'énergie dégagée par les différentes réactions d'interaction ( $\Delta H_n$ ) soit une fonction décroissante du nombre d'ARN<sub>t</sub><sup>Sec</sup> déjà présents sur le décimère, tout comme la constante de la réaction de liaison pour un état donné ( $k_n$ ).

La programmation, grâce au logiciel COPASI,<sup>[II.52]</sup> d'une modélisation de l'interaction SelA / ARN<sub>t</sub><sup>Sec</sup> a permis de générer un modèle correspondant aux données expérimentales de microcalorimétrie, dans le but de les interpréter quantitativement. Ce modèle, encore en développement, ne correspond pas aux possibilités expérimentales. Son amélioration et l'ajout de données expérimentales supplémentaires permettraient d'obtenir une simulation précise du comportement de cette interaction SelA / ARN<sub>t</sub><sup>Sec</sup> non acylé, et les valeurs de ses paramètres cinétiques et thermodynamiques.

Le système complexe permettant la formation de la sélénocystéine par réaction du séryl-ARN<sub>t</sub><sup>Sec</sup> avec SelA serait alors en partie dévoilé. En effet, le mécanisme de reconnaissance de l'ARN<sub>t</sub><sup>Sec</sup> et du séryl-ARN<sub>t</sub><sup>Sec</sup> au niveau de leur squelette ribose-phosphate doit posséder des paramètres cinétiques et thermodynamiques très similaires. Cela permettrait une meilleure compréhension de ce système, dont la complexité ne permet pas actuellement d'en proposer un modèle.

## 5.2 Étude structurale théorique sur SelA

L'objectif principal de mon travail de doctorat était l'obtention de la structure cristallographique de SelA, avec ou sans son partenaire ARN<sub>t</sub><sup>Sec</sup>.

Même si ce but n'a pas été totalement atteint en lui-même, les réflexions suscitées par l'étude du fonctionnement de l'enzyme peuvent apporter un point de vue intéressant sur les connaissances déjà disponibles concernant la structure de SelA. Comme ces connaissances existantes ne découlaient que des clichés de cryomicroscopie électronique, il m'a semblé utile de préciser notre idée de la structure de SelA grâce aux outils bio-informatiques de prédiction de structures secondaires disponibles.

Ces logiciels, généralement présents sur des serveurs de calcul académiques, en libre accès, ont beaucoup évolué et permettent aujourd'hui de disposer de prédiction de bonne qualité, dans le cas de protéines appartenant à des familles enzymatiques communes.

La sélénocystéine synthase possède une structure particulière, mais appartient à la grande famille des enzymes PLP-dépendantes de «Fold-Type I». Il est donc probable que la prédiction puisse aboutir à un résultat sensé. Pour en être certain, trois types différents de logiciels ont été utilisés pour prédire une structure secondaire à partir de la séquence primaire de SelA d'*Escherichia coli* et SelA de *Moorella thermoacetica*, puis les prédictions ont été confrontées pour vérifier si leurs caractéristiques faisait consensus.

### 5.2.1 Calculs de prédictions des structures secondaires de Sela

Les serveurs de prédiction de structure qui ont été utilisés, sont Phyre, Psipred et NPS@.

#### Phyre<sup>[II.54]</sup>

PHYRE est l'acronyme pour "Protein Homology/analogY Recognition Engine", un programme de reconstruction de structure secondaire par analogie de la séquence cible avec celles des protéines dont la structure est connue et présente dans la PDB.

Ce programme a été mis au point par le groupe du professeur Michael J.E. Sternberg, du centre de bio-informatique de l'Imperial College de Londres, et peut être utilisé sur leur serveur à l'adresse <http://www.sbg.bio.ic.ac.uk/phyre/index.cgi>.

Le principe de son fonctionnement est basé sur une première prédiction algorithmique de la structure secondaire de la séquence cible. La comparaison des profils enzymatiques (générés par Psi-Blast) de la séquence cibles et des séquences des protéines de structure connue permet alors de trouver des analogues, lesquels seront utilisés pour générer un alignement, à la fois des séquences et des structures secondaires communes.

Un modèle est alors reconstruit à partir des coordonnées de la structure résolue de l'analogue. Lorsqu'une différence importante de structure, ou un ajout / une délétion de séquence, sont présents entre le modèle et la cible le programme tente de générer la structure appropriée grâce à une librairie de structures et boucles existantes.

#### Psipred<sup>[II.55–II.58]</sup>

PSIPRED signifie "Protein Structure Prediction Server". L'algorithme de base proposé par le serveur effectue une prédiction calculatoire de la structure secondaire de la séquence cible à partir d'alignements par morceaux avec des protéines de structures connues. Pour augmenter la qualité des résultats obtenus, cet algorithme peut être combiné à la méthode "mGenTHREADER", qui fonctionne sur un principe similaire à Phyre. Son originalité réside dans l'utilisation préliminaire de Psipred. Si aucune structure n'est reconstruite par ce programme, on obtient des alignements de profil de structures de la prédiction avec une grande quantité d'analogues.

L'adresse de ce serveur est <http://bioinf.cs.ucl.ac.uk/psipred/>.

Ce programme est développé par le groupe du professeur David Jones, dans le département des sciences informatiques de l'University College de Londres.

#### NPS@<sup>[II.59]</sup>

NPS@ signifie "Network Protein Sequence @nalysis". Il s'agit d'un serveur français basé à l'Institut de Biologie et Chimie des Protéines (IBPC), du Pôle bioinformatique lyonnais de Gerland, développé par Gilbert Deléage.

Ce serveur permet un calcul de structure par 12 algorithmes différents, ainsi que la création d'une prédiction consensus.

Il est disponible à l'adresse <http://npsa-pbil.ibcp.fr> .

### 5.2.2 Comparaisons structurales avec les protéines homologues

Les prédictions des structures secondaires de SelA d'*E. coli* et de *M. thermoacetica* effectuées sur les trois serveurs précédents ont donné des résultats cohérents, en particulier au niveau de la région coeur, proche du site actif portant le PLP (voir § 2.3.2.2, page 92).

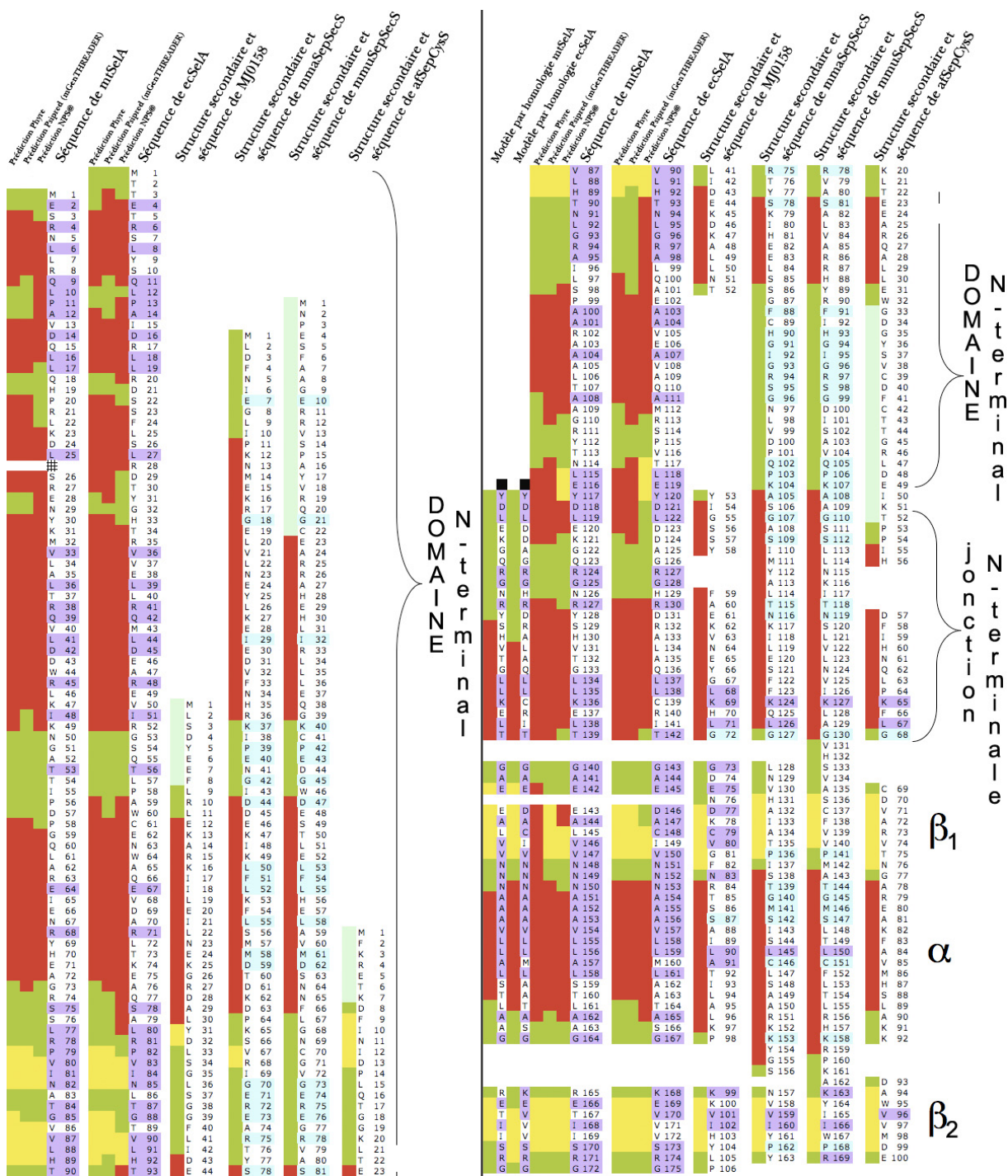
Les structures prédites m'ont permis de construire un alignement des séquences de SelA d'*E. coli* et de SelA de *M. thermoacetica* avec les protéines analogues suivantes, dont la structure est déjà résolue : la "SelA-like" MJ0158 de *M. jannaschii*,<sup>[II.1]</sup> la SepSecS de *Methanococcus maripaludis*<sup>[II.4]</sup> et de *Mus musculus*,<sup>[II.3]</sup> et enfin la SepCysS d'*Archaeoglobus fulgidus*.<sup>[II.43]</sup>

L'alignement en question est présenté sur les figures 5.2 à 5.4, page 213 à 215. La construction d'un tel alignement ne peut pas être effectuée grâce à une simple comparaison de séquences, ces protéines étant liées par trop peu d'identité de séquence. Les positions des résidus conservés sont tout de même soulignées par les couleurs mauve, pour les séquences de SelA, et bleue, pour les séquences de SepSecS. Il est à noter que la "SelA-like" MJ0158 est la protéine analogue qui possède la plus grande identité de séquence avec SelA.

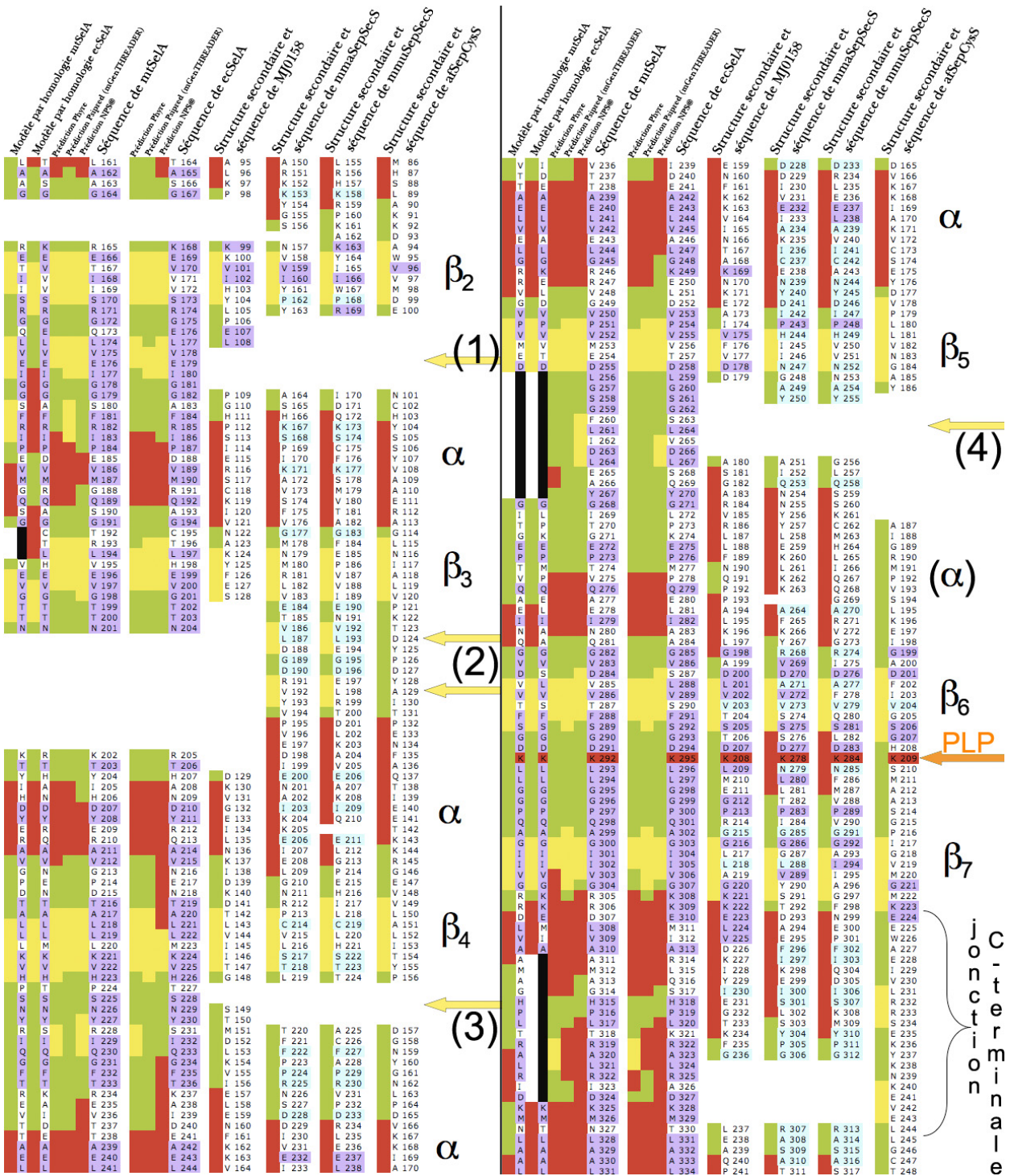
Malgré des séquences éloignées, toutes ces protéines possèdent un pattern commun :

- le domaine N-terminal est majoritairement constitué d'hélice  $\alpha$  ;
- la jonction entre ce domaine et le coeur caractéristique des enzymes de «Fold-Type I» est constituée d'une hélice  $\alpha$  d'une dizaine d'acides aminés ;
- le coeur catalytique suit l'organisation déjà décrite au § 2.3.2.2, page 92 : 7 brins  $\beta$  forment un feuillet pris en sandwich entre des hélices  $\alpha$  (4 pour la SepCysS, et 5 pour les autres enzyme) ;
- le pyridoxal-5'-phosphate est porté à une position identique sur les 6 protéines, sur la boucle joignant les brins  $\beta_6$  et  $\beta_7$  ;
- la jonction entre le coeur et la partie C-terminale est constituée d'une hélice  $\alpha$  pour toutes les protéines, à l'exception de la SepCysS, pour qui cette jonction est constituée de deux brins  $\beta$  ;
- la partie C-terminale de la protéine est caractérisée par une très grande hélice  $\alpha$  en début de sa séquence(jusqu'à 44 acides aminés pour les SepSecS), ce qui donne à toutes les enzymes présentées ici leur forme de T plus ou moins marquée ;
- le reste de la partie C-terminale présente une alternance entre des régions structurées en hélice  $\alpha$  et en brins  $\beta$ , qui permet généralement la formation d'un petit feuillet  $\beta$  dans le domaine ne portant pas le site catalytique.

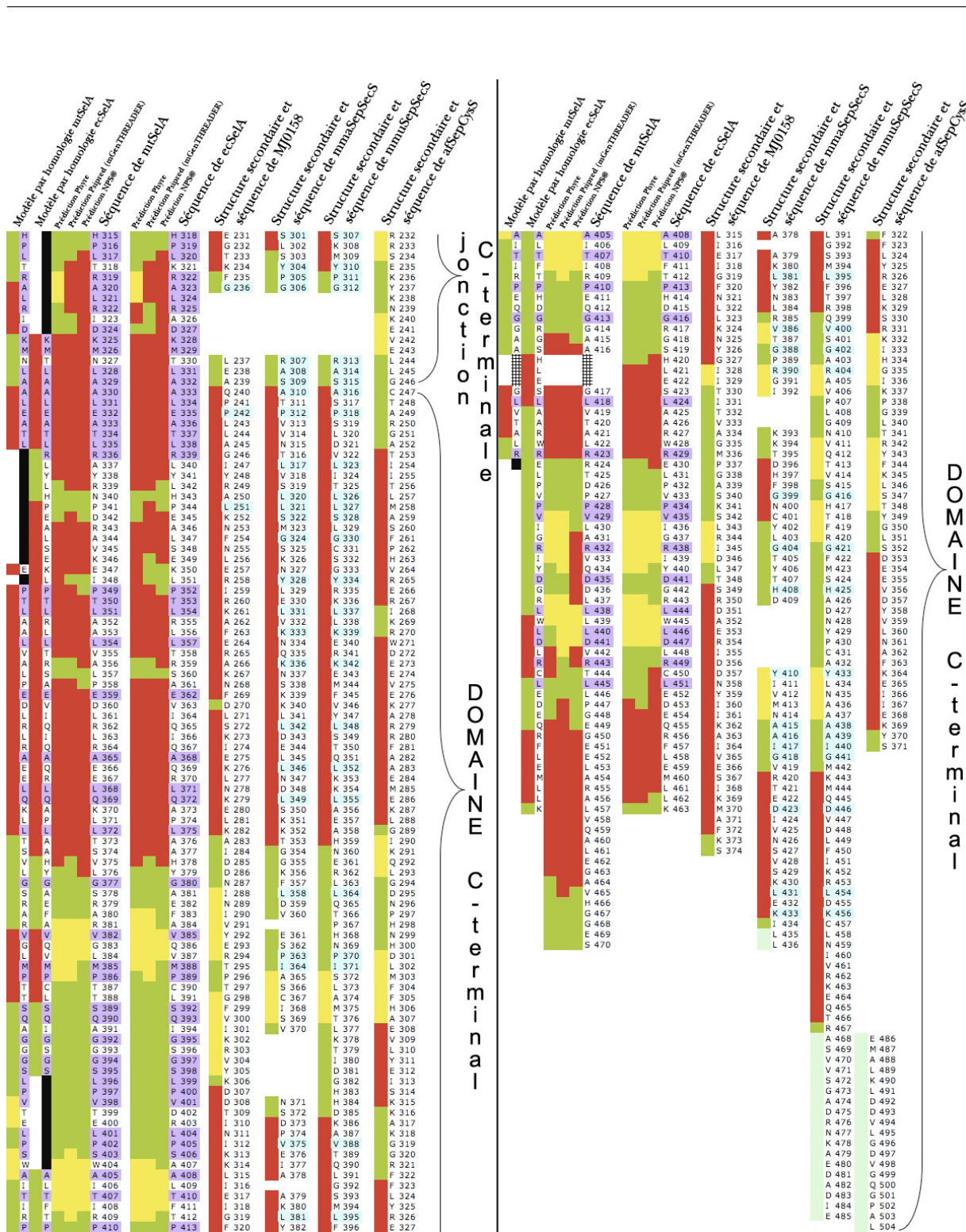
Pour visualiser toutes ces caractéristiques, et comparer ces protéines à d'autres enzymes PLP-dépendantes, un petit catalogue des structures en question est visible en annexe A.6, page 235.



**Figure 5.2** – Premier tiers de l'alignement des structures secondaires des protéines Sela, MJ0158, SepSecS et SepCysS. Les hélices  $\alpha$  sont représentées en rouge, les brins  $\beta$  en jaune et les régions non-structurées en vert ; le vert clair situe les régions absentes de la densité électronique des structures résolues.



**Figure 5.3** – Deuxième tiers de l’alignement des structures secondaires des protéines SclA, MJ0158, SepSecS et SepCysS. Les hélices  $\alpha$  sont représentées en rouge, les brins  $\beta$  en jaune et les régions non-structurées en vert. Les flèches numérotées de (1) à (4) pointent des particularités de l’alignement décrites dans le texte page 216.



**Figure 5.4** – Dernier tiers de l'alignement des structures secondaires des protéines Sela, MJ0158, SepSecS et SepCysS. Les hélices  $\alpha$  sont représentées en rouge, les brins  $\beta$  en jaune et les régions non-structurées en vert ; le vert clair situe les régions absentes de la densité électronique des structures résolues.



À coté de ces ressemblances entre les structures secondaires présentées, de grandes différences sont visibles et différencie ces protéines.

Ainsi, les flèches (1), (3) et (4) de la figure 5.3, page 214, montrent les régions pour lesquelles SelA possède une séquence plus longue que les autres protéines de l'alignement. Cette insertion est d'ailleurs caractérisée par la présence possible de structures secondaires additionnelles, sous la forme de brins  $\beta$ . La présence d'un brin  $\beta$  à la suite du brin  $\beta_2$  (flèche (1)) est d'ailleurs prédite par l'unanimité des serveurs utilisés.

Les flèches (2) de la figure 5.3, page 214, désignent à la fois une zone de délétion de la séquence des deux SelA et de MJ0158, et la présence d'une structure secondaire particulière sur les SepSecS. Ces dernières présentent deux brins  $\beta$  surnuméraires, bien visibles sur la structure en (B) de la figure 5.5, page suivante. Leur importance dans le fonctionnement des SepSecS n'est pas connue.

Les différences et les ressemblances entre ces protéines analogues permettent tout de même de confirmer une tendance déjà remarquée au § 2.3.5, page 98 : de ces trois enzymes analogues, SepCysS est celle qui est structurellement la plus éloignée de SelA ; et la "SelA-like" MJ0158 est celle qui exhibe le plus de similitude avec SelA, malgré sa taille bien moindre.

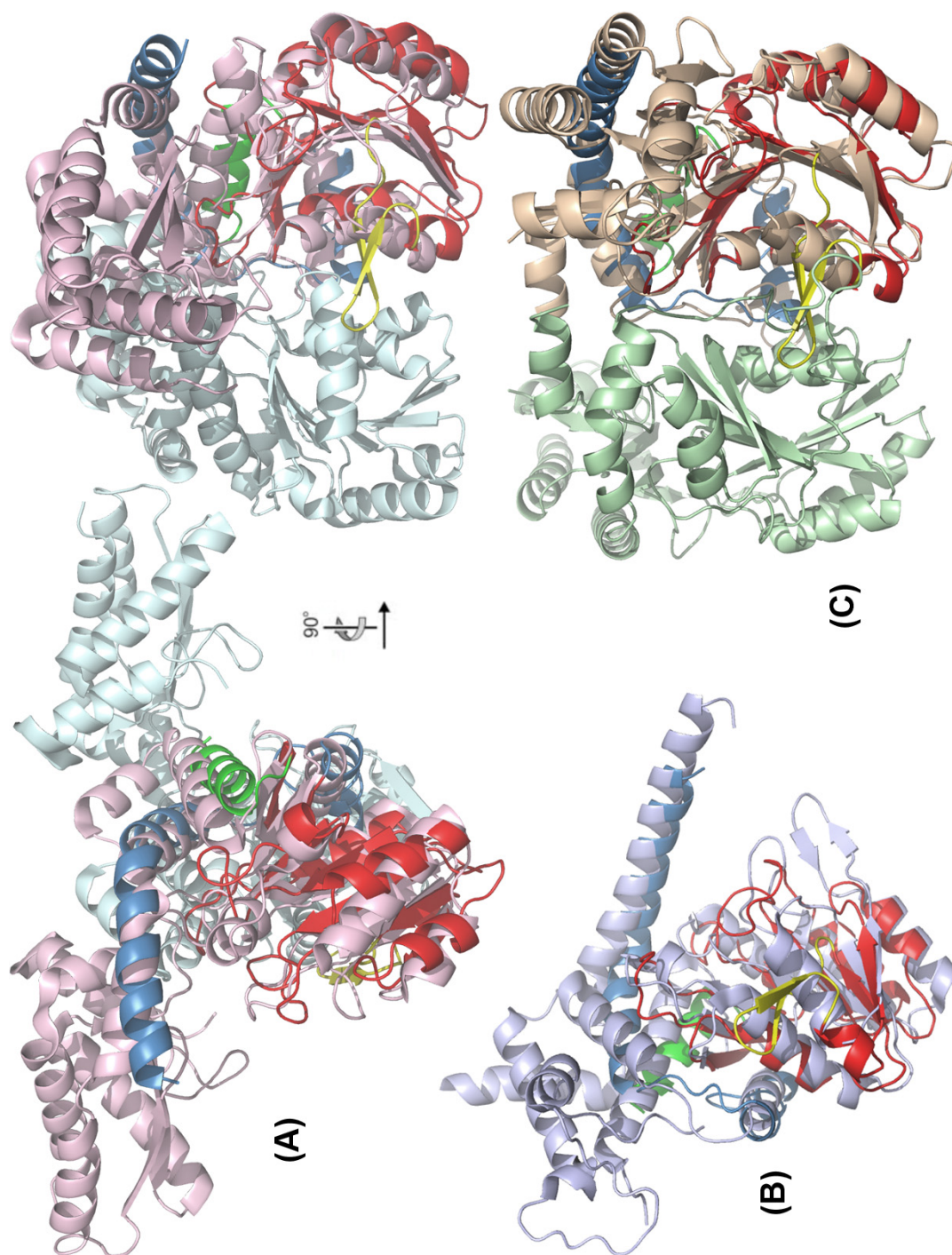
### Analyse des structures reconstruites de Sela

Le programme de prédiction de structure Phyre génère deux résultats différents : d'un coté une prédiction de structure algorithmique qui est présentée dans les colonnes notées "Phyre" sur les alignements précédents, et de l'autre, une reconstruction d'une structure tridimensionnelle de la séquence cible — ici SelA — à partir du plus proche analogue dont la structure est connue et d'une librairie de motifs structurels connus en cas de divergence par rapport à cet analogue.

Les calculs effectués aboutissent à deux fichiers PDB contenant les coordonnées des modèles pour SelA d'*E. coli* et de *M. thermoacetica*. Les structures secondaires de ces modèles, légèrement différentes des autres prédictions, sont présentées dans les colonnes notées "Modèles par homologie" sur la gauche des alignements précédents.

Les parties N-terminale et C-terminale de ces protéines étant très différentes des structures connues, la reconstruction du modèle n'a pris en compte que les zones de jonction et le coeur des protéines SelA. On peut également remarquer, sur l'alignement des pages précédentes, des zones noires indiquant une délétion dans le modèle obtenu, sans doute pour les mêmes raisons.

La figure 5.5, page suivante présente un alignement tridimensionnelle entre la structure du modèle généré pour SelA de *M. thermoacetica*, et celles des trois autres protéines analogues. Les structures sont alignées par rapport au feuillet  $\beta$  de leur domaine coeur.

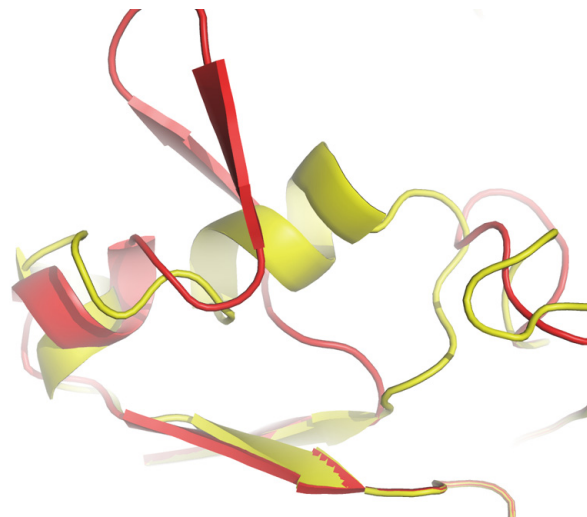


**Figure 5.5** – Alignement des structures de la prédiction de SelA de *M. thermoacetica* avec : la “SelA-like” MJ0158 en (A), la SepSecS de *Mus musculus* en (B) et la SepCysS en (C). Le modèle est présenté dans le code couleur tricolore : vert = N-ter, rouge = Coeur, et bleu = C-ter. En jaune est indiquée la structure supplémentaire de SelA en brin  $\beta$ . En (B), seul la grande hélice de la partie C-ter de SepSecS est représentée, pour préserver la clarté de la figure.

Dans la figure 5.5 précédente, seul le modèle obtenu pour SelA de *M. thermoacetica* est pris comme point de comparaison. SelA *E. coli* n'a pas été utilisée parce que le modèle obtenu est très similaire à celui de *Moorella thermoacetica*. Le programme Phyre a en effet utilisé le même analogue comme point de départ de la reconstruction de structure : la L-méthionine-lyase de *Citrobacter Freundii* (code d'accès PDB : 1Y4I, voir annexe A.6, page 235). Les deux modèles ne diffèrent qu'au niveau des zones qui ne correspondent pas avec cet analogue, et pour lesquelles le programme a utilisé sa librairie de motifs structuraux connus.

Il est d'ailleurs intéressant de noter que, parmi les zones différentes dans les prédictions de SelA de *E. coli* et de *M. thermoacetica*, se trouve la zone indiquée par la flèche (1), zone pour laquelle une structure secondaire, sous forme de brin  $\beta$ , est décrite par tous les programmes de prédiction.

Pourtant, dans le modèle reconstruit d'*Escherichia coli*, qui, lui, est obtenu de façon plus séquence-dépendante et en fonction des données de la librairie utilisée, cette structure est reconstruite sous la forme d'une hélice  $\alpha$  (voir le figure 5.6).



**Figure 5.6** – Comparaison de la zone indiquée par la flèche (1), dans les prédictions des structures de SelA de *M. thermoacetica* (en rouge) et d'*E. coli* (en jaune).

Que ce soit la majorité ou le rapport minoritaire qui soit le plus proche de la vérité importe peu. Ces 8 prédictions de structures secondaires (dont deux sont des reconstructions de structures tridimensionnelles) mettent en exergue la présence, chez SelA, de boucles le taille plus importante qu'à la normale dans une zone spatiale restreinte, sur un des cotés du domaine coeur (c-à-d. sur les boucles après les brins  $\beta_2$ ,  $\beta_4$  et  $\beta_5$  : la figure 2.6, page 93, permet de repérer ces boucles dans l'organisation spatiale du coeur de la protéine PLP-dépendante).

L'organisation oligomérique de Sela pourrait donc passer par une interaction inter-dimère au niveau de ces boucles allongées. Cette possibilité peut être confirmée par l'étude des configurations quaternaires existant chez les protéines PLP-dépendantes.

### Banque de protéines PLP-dépendantes analogues

En effet, l'utilisation des programmes Psipred et Phyre a permis de constituer une banque d'enzyme PLP-dépendante dont l'organisation des structures secondaires est analogue à celle de Sela : ces protéines, qui n'auraient pu être identifiées par une simple recherche par comparaison de séquence, ont des identités de séquences allant de 5% à 20% avec les protéines Sela, lorsque l'alignement des structures secondaires est effectué.

Celles qui possèdent le plus d'identité de séquence avec Sela d'*E. coli* et/ou Sela de *M. thermoacetica* sont représentées dans le catalogue de structure de l'annexe A.6, page 235.

On peut y voir une majorité de protéines dimériques, mais également quelques tétramères. Un tétramère en particulier a attiré mon attention : la thréonine aldolase (n° d'accès PDB : 1M6S, voir la figure A.12, page 240), dont l'organisation tétramérique n'est pas conventionnelle.

Toutes les enzymes PLP-dépendantes de «Fold-Type I» tétramériques vues jusqu'à présent exhibaient une liaison entre dimères au niveau de leur domaine formé des extrémités N-terminale et C-terminale — comme les SepSecS (n° PDB 2Z67 et 3BC8), la L-méthionine-lyase (n° PDB 1Y4I), la O-acétyl-homosérine sulphydrylase(2CB1) et la cystathionine  $\gamma$ -lyase (1CS1), voir les figures en annexe A.6, page 235.

Le tétramère de la thréonine aldolase de *Thermotoga maritima* est composé de deux dimères liés par leur domaine coeur, grâce à une boucle de taille importante (figure A.12, page 240).

### 5.2.3 Vers un modèle structural de Sela ?

Comme la possibilité d'une organisation oligomérique basée sur une liaison de boucles du domaine coeur est envisageable, le site de l'interaction inter-dimère de Sela prend sans doute forme dans les boucles présentes juste après les brins  $\beta 2$ ,  $\beta 4$  et  $\beta 5$  constitutifs du coeur PLP (flèches 1, 3 et 4).

L'organisation décamerique de Sela est le facteur générateur de complexité, pour la formation d'une population homogène de complexes et pour l'étude des interactions enzymatiques entre Sela et l'ARN<sub>t</sub><sup>Sec</sup>.

Une nouvelle piste est envisageable bien que complexe à mettre en oeuvre : la conception par génie génétique d'une protéine Sela chimère pourrait faciliter l'étude de la biosynthèse de la sélénocystéine chez les eubactéries.

Cette protéine SelA chimère, serait caractérisée par la séquence de SelA<sub>sauvage</sub>, sauf au niveau des boucles, précédemment identifiées, par lesquelles pourrait se faire l'interaction inter-dimère. Ces boucles pourraient être remplacées par celle de SepSecS ou de MJ0158.

Si cette protéine est soluble, reste sous forme de dimère, et conserve une activité sélénocystéine synthase, cela prouverait que l'organisation décamerique de SelA n'existe pas dans un but réactionnel. La formation du complexe "1 pour 1" entre cette chimère et le séryl-ARN<sub>t</sub><sup>Sec</sup> serait simplifiée, et l'étude de leur interaction facilitée. Les données obtenues soulèveraient le voile sur la première étape de complexation d'un ARN<sub>t</sub><sup>Sec</sup> sur la sélénocystéine synthase décamerique.

Dans le cas contraire, où cette chimère serait soluble et dimérique mais ne présenterait pas d'activité sélénocystéine synthase, l'hypothèse de l'implication de l'organisation oligomérique dans la fonction de SelA serait confortée.

### 5.3 Perspectives

Ce projet de recherche sur l'acteur de la biosynthèse de la sélénocystéine SelA a pu montrer son intérêt : le cas complexe d'interaction qu'il représente est un défi que j'ai eu plaisir à relever pendant ce travail de doctorat.

Des améliorations peuvent néanmoins être tentées pour faciliter l'accès à des informations d'intérêt sur ce système, en particulier en modifiant notre protéine cible. Les techniques de purification mises au point pour les constructions protéiques de SelA que nous avons étudiées sont transposables à un cas similaire.

Les outils et les techniques développés lors de ce travail peuvent également prétendre à une portée plus large, et pourraient faire l'objet d'une validation par leur application sur d'autres objets d'études. Je pense en particulier aux outils de simulation d'isotherme de liaison qui rendrait possible, après son amélioration, l'interprétation des données obtenues par ITC pour des systèmes complexes.

De même la technique de SHAPE appliquée à un complexe protéine / ARN pourrait faire l'objet de tests sur d'autres complexes, et ainsi permettre de valider cette méthode.

Bien sûr, les avancées effectuées sur la formation du complexe entre SelA et l'ARN<sub>t</sub><sup>Sec</sup>, jusqu'à sa cristallisation, seraient à prolonger pour pouvoir obtenir, finalement, la structure cristallographique de ces macromolécules.

# Références bibliographiques de la Partie II

- [1] KAISER J. T. ; GROMADSKI K. ; ROTHER M. ; ENGELHARDT H. ; RODNINA M. V. ; WAHL M. C. : Structural and functional investigation of a putative archaeal selenocysteine synthase. *Biochemistry*, 44(40) : p13315–13327, Oct 2005.
- [2] YUAN J. ; PALIOURA S. ; SALAZAR J. C. ; SU D. ; O'DONOGHUE P. ; HOHN M. J. ; CARDOSO A. M. ; WHITMAN W. B. ; SÖLL D. : RNA-dependent conversion of phosphoserine forms selenocysteine in eukaryotes and archaea. *Proceedings of the National Academy of Sciences of the United States of America*, 103(50) : p18923–18927, Dec 2006.
- [3] GANICHKIN O. M. ; XU X.-M. ; CARLSON B. A. ; MIX H. ; HATFIELD D. L. ; GLADYSHEV V. N. ; WAHL M. C. : Structure and catalytic mechanism of eukaryotic selenocysteine synthase. *The Journal of biological chemistry*, 283(9) : p5849–5865, Feb 2008.
- [4] ARAISO Y. ; PALIOURA S. ; ISHITANI R. ; SHERRER R. L. ; O'DONOGHUE P. ; YUAN J. ; OSHIKANE H. ; DOMAE N. ; DEFRANCO J. ; SÖLL D. ; NUREKI O. : Structural insights into RNA-dependent eukaryal and archaeal selenocysteine formation. *Nucleic acids research*, 36(4) : p1187–1199, Mar 2008.
- [5] FISCHER N. ; PALESKAVA A. ; GROMADSKI K. B. ; KONEVEGA A. L. ; WAHL M. C. ; STARK H. ; RODNINA M. V. : Towards understanding selenocysteine incorporation into bacterial proteins. *Biological chemistry*, 388(10) : p1061–1067, Oct 2007.
- [6] LEINFELDER W. ; STADTMAN T. C. ; BÖCK A. : Occurrence in vivo of selenocysteyl-tRNA(SERUCA) in Escherichia coli. Effect of sel mutations. *The Journal of biological chemistry*, 264(17) : p9720–9723, Jun 1989.
- [7] SCHÖN A. ; BÖCK A. ; OTT G. ; SPRINZL M. ; SÖLL D. : The selenocysteine-inserting opal suppressor serine tRNA from E. coli is highly unusual in structure and modification. *Nucleic acids research*, 17(18) : p7159–7165, Sep 1989.
- [8] FORCHHAMMER K. ; BÖCK A. : Selenocysteine synthase from Escherichia coli. Analysis of the reaction sequence. *The Journal of biological chemistry*, 266(10) : p6324–6328, Apr 1991.
- [9] ENGELHARDT H. ; FORCHHAMMER K. ; MÜLLER S. ; GOLDIE K. N. ; BÖCK A. : Structure of selenocysteine synthase from Escherichia coli and location of tRNA in the seryl-tRNA(sec)-enzyme complex. *Molecular microbiology*, 6(23) : p3461–3467, Dec 1992.
- [10] THANBICHLER M. ; BÖCK A. : Selenoprotein biosynthesis : purification and assay of components involved in selenocysteine biosynthesis and insertion in Escherichia coli. *Methods in enzymology*, 347 : p3–16, 2002.
- [11] TORMAY P. ; WILTING R. ; HEIDER J. ; BÖCK A. : Genes coding for the selenocysteine-inserting tRNA species from Desulfomicrobium baculatum and Clostridium thermoaceticum : structural and evolutionary implications. *Journal of bacteriology*, 176(5) : p1268–1274, Mar 1994.
- [12] MATTHAEI J. H. ; VOIGT H. P. ; HELLER G. ; NETH R. ; SCHÖCH G. ; KÜBLER H. ; AMELUNXEN F. ; SANDER G. ; PARMEGGIANI A. : Specific interactions of ribosomes in decoding. *Cold Spring Harbor symposia on quantitative biology*, 31 : p25–38, xx 1966.

- [13] BARDIN C. ; LEROY J. L. : The formation pathway of tetramolecular G-quadruplexes. *Nucleic acids research*, 36(2) : p477–488, Feb 2008.
- [14] FORCHHAMMER K. ; LEINFELDER W. ; BOESMILLER K. ; VEPREK B. ; BÖCK A. : Selenocysteine synthase from *Escherichia coli*. Nucleotide sequence of the gene (*selA*) and purification of the protein. *The Journal of biological chemistry*, 266(10) : p6318–6323, Apr 1991.
- [15] MEINNEL T. ; MECHULAM Y. ; FAYAT G. : Fast purification of a functional elongator tRNA<sup>Met</sup> expressed from a synthetic gene in vivo. *Nucleic acids research*, 16(16) : p8095–8096, Aug 1988.
- [16] MEINNEL T. ; BLANQUET S. : Maturation of pre-tRNA(fMet) by *Escherichia coli* RNase P is specified by a guanosine of the 5'-flanking sequence. *The Journal of biological chemistry*, 270(26) : p15908–15914, Jun 1995.
- [17] JÜNEMANN R. ; WADZACK J. ; TRIANA-ALONSO F. J. ; BITTNER J. U. ; CAILLET J. ; MEINNEL T. ; VANATALU K. ; NIERHAUS K. H. : In vivo deuteration of transfer RNAs : overexpression and large-scale purification of deuterated specific tRNAs. *Nucleic acids research*, 24(5) : p907–913, Mar 1996.
- [18] TISNÉ C. ; RIGOURD M. ; MARQUET R. ; EHRESMANN C. ; DARDEL F. : NMR and biochemical characterization of recombinant human tRNA(Lys)<sub>3</sub> expressed in *Escherichia coli* : identification of posttranscriptional nucleotide modifications required for efficient initiation of HIV-1 reverse transcription. *RNA*, 6(10) : p1403–1412, Oct 2000.
- [19] KOTHE U. ; PALESKAVA A. ; KONEVEGA A. L. ; RODNINA M. V. : Single-step purification of specific tRNAs by hydrophobic tagging. *Analytical biochemistry*, 356(1) : p148–150, Sep 2006.
- [20] SOLER N. ; FOURMY D. ; YOSHIZAWA S. : Crystallization and preliminary X-ray analysis of the mRNA-binding domain of elongation factor SelB from *Escherichia coli* in complex with RNA. *Acta crystallographica. Section F, Structural biology and crystallization communications*, 63(Pt 5) : p419–421, May 2007.
- [21] SOLER N. ; FOURMY D. ; YOSHIZAWA S. : Structural insight into a molecular switch in tandem winged-helix motifs from elongation factor SelB. *Journal of molecular biology*, 370(4) : p728–741, Jul 2007.
- [22] YOSHIZAWA S. ; RASUBALA L. ; OSE T. ; KOHDA D. ; FOURMY D. ; MAENAKA K. : Structural basis for mRNA recognition by elongation factor SelB. *Nature structural & molecular biology*, 12(2) : p198–203, Feb 2005.
- [23] SOLER N. : *Etude structurale de la protéine SelB, impliquée dans l'incorporation des sélénocystéines lors de la synthèse protéique bactérienne*. 216 pages. Thèse de Doctorat, Biochimie, Biologie structurale, Université Paris 11, Jul 2007.
- [24] OSE T. ; SOLER N. ; RASUBALA L. ; KUROKI K. ; KOHDA D. ; FOURMY D. ; YOSHIZAWA S. ; MAENAKA K. : Structural basis for dynamic interdomain movement and RNA recognition of the selenocysteine-specific elongation factor SelB. *Structure*, 15(5) : p577–586, May 2007.
- [25] SOLER N. ; FOURMY D. ; YOSHIZAWA S. : Molecular switch in tandem winged-helix motifs of elongation factor SelB. *Nucleic acids symposium series*, (51) : p377–378, 2007.
- [26] HUSSON C. : *Reconnaissance et modification de l'ARN ribosomique par la méthyltransférase RlmAII à l'origine de la résistance bactérienne à l'antibiotique macrolide tylosine*. Thèse de Doctorat, Biochimie, Biologie structurale, Université Paris 11, Dec 2007.
- [27] LEBARS I. ; HUSSON C. ; YOSHIZAWA S. ; DOUTHWAITE S. ; FOURMY D. : Recognition elements in rRNA for the tylosin resistance methyltransferase RlmA(II). *Journal of molecular biology*, 372(2) : p525–534, Sep 2007.
- [28] FORCHHAMMER K. ; BOESMILLER K. ; BÖCK A. : The function of selenocysteine synthase and SELB in the synthesis and incorporation of selenocysteine. *Biochimie*, 73(12) : p1481–1486, Dec 1991.
- [29] MIZUTANI T. ; KURATA H. ; YAMADA K. ; TOTSUKA T. : Some properties of murine selenocysteine synthase. *The Biochemical journal*, 284(Pt 3) : p827–834, Jun 1992.

- 
- [30] BARON C.; STURCHLER C.; WU X. Q.; GROSS H. J.; KROL A.; BÖCK A.: Eukaryotic selenocysteine inserting tRNA species support selenoprotein synthesis in *Escherichia coli*. *Nucleic acids research*, 22(12) : p2228–2233, Jun 1994.
- [31] TORMAY P.; WILTING R.; LOTTSPREICH F.; MEHTA P. K.; CHRISTEN P.; BÖCK A.: Bacterial selenocysteine synthase—structural and functional properties. *European journal of biochemistry / FEBS*, 254(3) : p655–661, Jun 1998.
- [32] MIZUTANI T.; KANAYA K.; TANABE K.: Selenophosphate as a substrate for mammalian selenocysteine synthase, its stability and toxicity. *BioFactors*, 9(1) : p27–36, 1999.
- [33] BORECKA B.; CHOJNOWSKI J.; CYPRYK M.; MICHALSKI J.; ZIELINSKA J.: Synthetic and mechanistic aspects of the reaction of trialkylsilyl halides with thio and seleno esters of phosphorus. *Journal of Organometallic Chemistry*, 171(1) : p17–34, 1979.
- [34] GLASS R. S.; SINGH W. P.; JUNG W.; VERES Z.; SCHOLZ T. D.; STADTMAN T. C.: Mono-selenophosphate : synthesis, characterization, and identity with the prokaryotic biological selenium donor, compound SePX. *Biochemistry*, 32(47) : p12555–12559, Nov 1993.
- [35] GLASS R. S.; STADTMAN T. C.: Selenophosphate. *Methods in enzymology*, 252 : p309–315, 1995.
- [36] BHARDWAJ R. K.; DAVIDSON R. S.: Oxidative desulphurisation and deselenation at pentavalent phosphorous by photogenerated peroxidic species. *Tetrahedron*, 43(19) : p4473–4479, 1987.
- [37] BARON C.; BÖCK A.: The length of the aminoacyl-acceptor stem of the selenocysteine-specific tRNA(Sec) of *Escherichia coli* is the determinant for binding to elongation factors SELB or Tu. *The Journal of biological chemistry*, 266(30) : p20375–20379, Oct 1991.
- [38] EHRESMANN C.; BAUDIN F.; MOUGEL M.; ROMBY P.; EBEL J. P.; EHRESMANN B.: Probing the structure of RNAs in solution. *Nucleic acids research*, 15(22) : p9109–9128, Nov 1987.
- [39] MERINO E. J.; WILKINSON K. A.; COUGHLAN J. L.; WEEKS K. M.: RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *Journal of the American Chemical Society*, 127(12) : p4223–4231, Mar 2005.
- [40] WILKINSON K. A.; MERINO E. J.; WEEKS K. M.: RNA SHAPE chemistry reveals nonhierarchical interactions dominate equilibrium structural transitions in tRNA(Asp) transcripts. *Journal of the American Chemical Society*, 127(13) : p4659–4667, Apr 2005.
- [41] WILKINSON K.; MERINO E.; WEEKS K.: Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) : quantitative RNA structure analysis at single nucleotide resolution. *Nature protocols*, 1(3) : p1610–1616, Nov 2006.
- [42] MORTIMER S. A.; WEEKS K. M.: A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *Journal of the American Chemical Society*, 129(14) : p4144–4145, Apr 2007.
- [43] FUKUNAGA R.; YOKOYAMA S.: Structural insights into the second step of RNA-dependent cysteine biosynthesis in archaea : crystal structure of Sep-tRNA :Cys-tRNA synthase from *Archaeoglobus fulgidus*. *Journal of molecular biology*, 370(1) : p128–141, Jun 2007.
- [44] PIERCE M. M.; RAMAN C. S.; NALL B. T.: Isothermal titration calorimetry of protein-protein interactions. *Methods*, 19(2) : p213–221, Oct 1999.
- [45] PRIVALOV G. P.; PRIVALOV P. L.: Problems and prospects in microcalorimetry of biological macromolecules. *Methods in enzymology*, 323 : p31–62, 2000.
- [46] PRIVALOV P. L.; DRAGAN A. I.: Microcalorimetry of biological macromolecules. *Biophysical chemistry*, 126(1-3) : p16–24, Mar 2007.
- [47] BUURMA N. J.; HAQ I.: Advances in the analysis of isothermal titration calorimetry data for ligand-DNA interactions. *Methods*, 42(2) : p162–172, Jun 2007.
- [48] HYRE D. E.; SPICER L. D.: Thermodynamic evaluation of binding interactions in the methionine repressor system of *Escherichia coli* using isothermal titration calorimetry. *Biochemistry*, 34(10) : p3212–3221, Mar 1995.



- [49] BRUZZESE F. J. ; CONNELLY P. R. : Allosteric properties of inosine monophosphate dehydrogenase revealed through the thermodynamics of binding of inosine 5'-monophosphate and mycophenolic acid. Temperature dependent heat capacity of binding as a signature of ligand-coupled conformational e. *Biochemistry*, 36(34) : p10428–10438, Aug 1997.
- [50] GOPAL B. ; SWAMINATHAN C. P. ; BHATTACHARYA S. ; BHATTACHARYA A. ; MURTHY M. R. ; SUROLIA A. : Thermodynamics of metal ion binding and denaturation of a calcium binding protein from *Entamoeba histolytica*. *Biochemistry*, 36(36) : p10910–10916, Sep 1997.
- [51] MENZE M. ; HELLMANN N. ; DECKER H. ; GRIESHABER M. : Binding of urate and caffeine to haemocyanin analysed by isothermal titration calorimetry. *The Journal of experimental biology*, 204(Pt 5) : p1033–1038, Mar 2001.
- [52] HOOPS S. ; SAHLE S. ; GAUGES R. ; LEE C. ; PAHLE J. ; SIMUS N. ; SINGHAL M. ; XU L. ; MENDES P. ; KUMMER U. : COPASI — a CComplex PATHway SIMulator. *Bioinformatics*, 22 : p3067–3074, 2006.
- [53] BARON C. ; HEIDER J. ; BÖCK A. : Mutagenesis of selC, the gene for the selenocysteine-inserting tRNA-species in *E. coli* : effects on in vivo function. *Nucleic acids research*, 18(23) : p6761–6766, Dec 1990.
- [54] BENNETT-LOVSEY R. ; HERBERT A. ; STERNBERG M. ; KELLEY L. : Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins : Structure, Function, Bioinformatics*, 70 : 3, 2008.
- [55] BRYSON K. ; MCGUFFIN L. ; MARSDEN R. ; WARD J. ; SODHI J. ; JONES D. : Protein structure prediction servers at University College London. *Nucleic acids research*, 33 : (Web Server issue) :W36-38, 2005.
- [56] JONES D. : Protein secondary structure prediction based on position-specific scoring matrices. *Journal of molecular biology*, 292 : p195–202, 1999.
- [57] MCGUFFIN L. ; JONES D. : Improvement of the GenTHREADER method for genomic fold recognition. *Bioinformatics*, 19 : p874–881, 2003.
- [58] DT J. : GenTHREADER : an efficient and reliable protein fold recognition method for genomic sequences. *Journal of molecular biology*, 287 : p797–815, 1999.
- [59] COMBET C. ; BLANCHET C. ; GEOURJON C. ; DELÉAGE G. : NPS@ : Network Protein Sequence Analysis. *TIBS*, 25(291) : 3 : p147–150, Mar 2000.

# Annexes



# Table des matières

---

<b>Annexe A</b>	<b>Annexes bibliographiques</b>	<b>229</b>
A.1	Arbre phylogénique de procaryotes séquencés . . . . .	229
A.2	Locus et opérons des gènes <i>selABCD</i> . . . . .	229
A.3	Phylogénie de Sela chez les eubactéries . . . . .	232
A.4	Les séquences protéiques de ecSela et de mtSela . . . . .	232
A.5	Arbres phylogéniques des enzymes PLP-dépendantes . . . . .	234
A.6	Structures d'enzymes PLP-dépendantes . . . . .	235
A.7	Structures des protéines décamériques connues . . . . .	241
<b>Annexe B</b>	<b>Matériels et Méthodes</b>	<b>245</b>
B.1	Fiches techniques . . . . .	245
B.1.1	Plasmide utilisé pour le clonage de <i>selC</i> . . . . .	245
B.1.2	Plasmide utilisé pour le clonage de Sela . . . . .	246
B.2	Protocoles expérimentaux . . . . .	247
B.2.1	PCR de multiplication de la matrice de transcription . . . . .	247
B.2.2	Transcription de <i>selC</i> en $\text{ARN}_t^{\text{Sec}}$ . . . . .	248
B.2.3	Charge en sérine de l' $\text{ARN}_t^{\text{Sec}}$ . . . . .	249
<b>Annexe C</b>	<b>La chimie du sélénium</b>	<b>251</b>

---



# Annexe A

## Annexes bibliographiques

### A.1 Arbre phylogénique de procaryotes séquencés

La reconstruction de cet arbre phylogénique (figure A.2, page suivante) utilise la comparaison entre la petite sous-unité du ribosome de ces organismes.<sup>[A.1]</sup> Sur cet arbre est reporté l'occurrence chez les organismes des caractères d'utilisations du sélénium sous forme de sélénocystéine (présence des gènes *selABC* et *selD*) et/ou de sélénouridine (présence des gènes *selD* et *ybbB*).

La branche basse de l'arbre regroupe les organismes appartenant au domaine des archées.

### A.2 Locus et opérons des gènes *selABCD*

Organisation en opéron des gènes des protéines nécessaires à la biosynthèse et à l'incorporation de la sélénocystéine, *selAB* et *selD*, chez *Escherichia coli*.<sup>[A.2]</sup> Un opéron est formé par *selAB* (leurs gènes se chevauchent). Un autre opéron est formé par *selD*, le gène encodant pour la topoisomérase III (*TopB*) et un gène *orf183* de fonction inconnue.

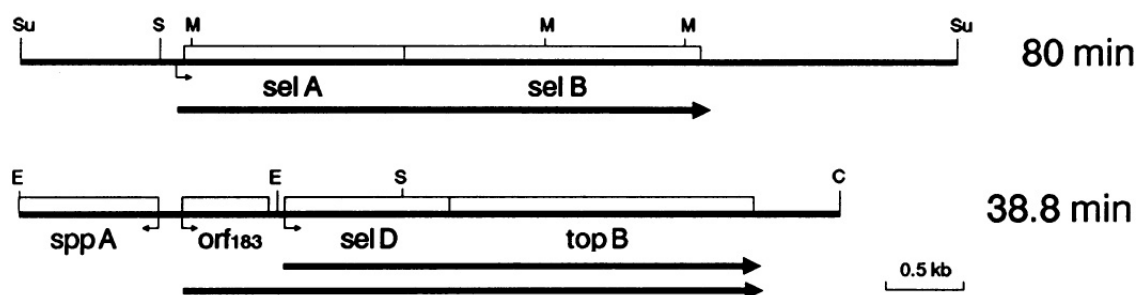
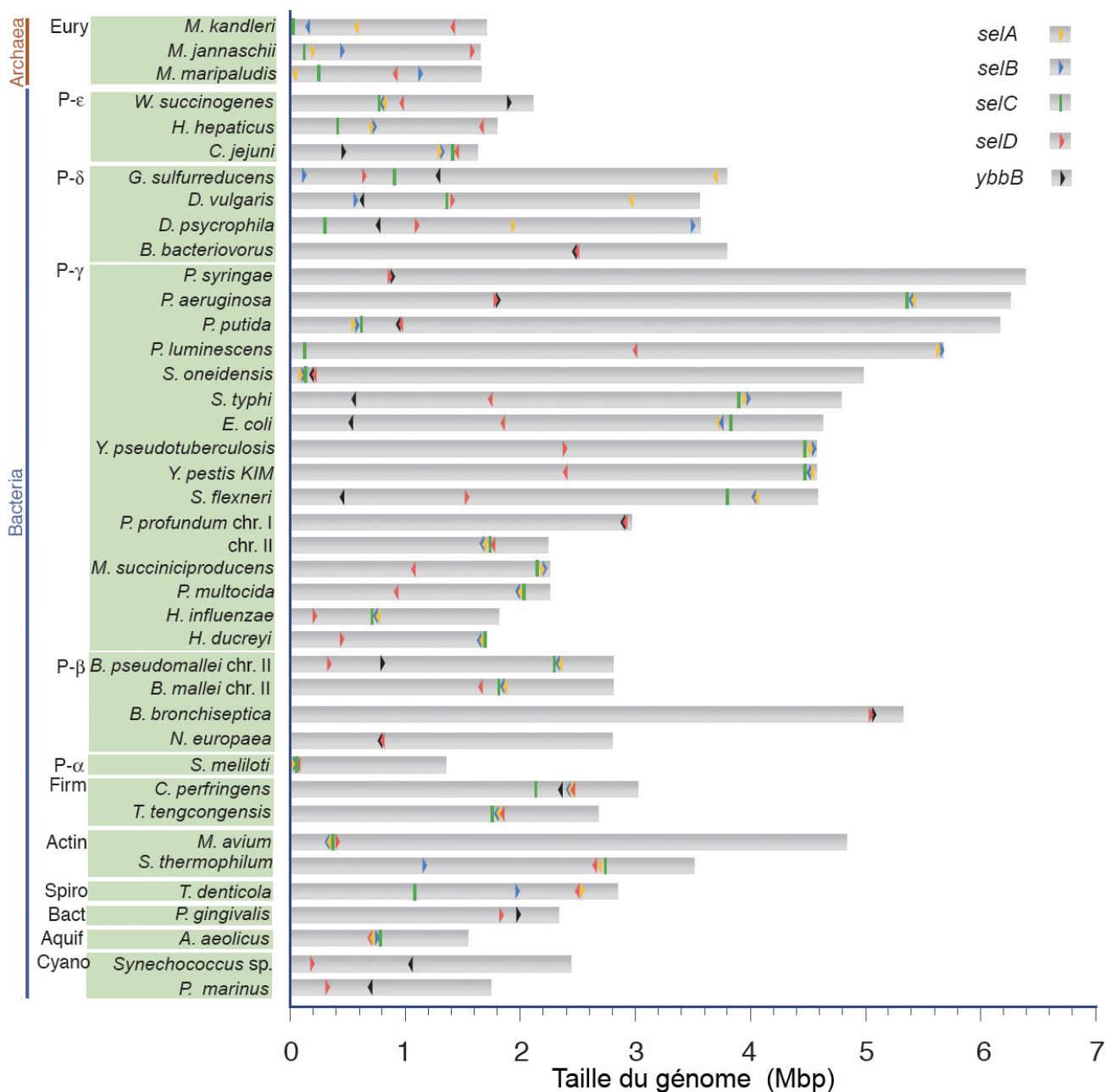


Figure A.1 – Locus des gènes *selAB* et *selD* sur le chromosome de *E. coli*.<sup>[A.2]</sup>

Les locus de ces gènes, ainsi que celui du gène *ybbB* nécessaire à la synthèse de la 2-sélénouridine (voir § 4, page 30), pour les eubactéries et archées séquencées en 2005, sont visibles figure A.3, page 231.<sup>[A.1]</sup>



Figure A.2 – Arbre phylogénique reconstruit, portant l'occurrence des gènes caractéristiques des utilisations du sélénium (Sec et Se<sup>2</sup>U) chez les procaryotes et archeaes séquencés en 2005.<sup>[A.1]</sup>

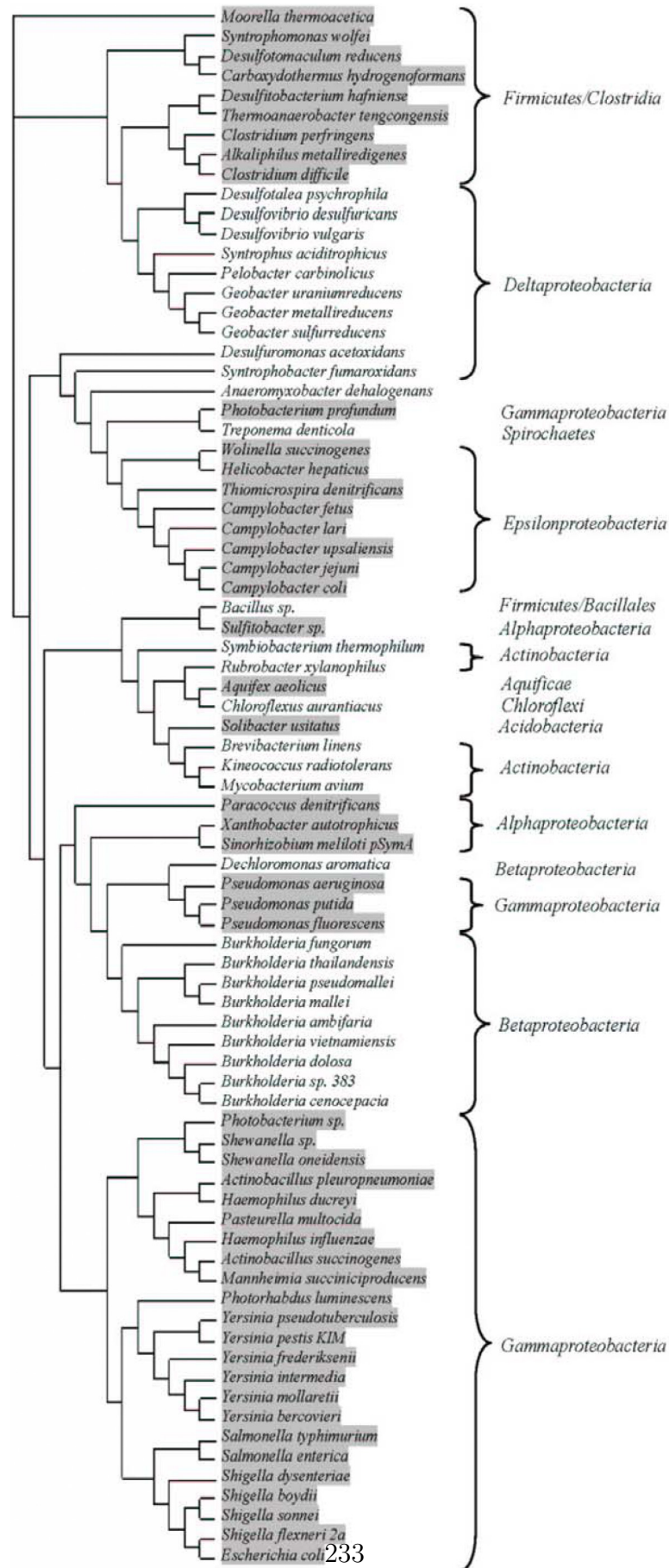


**Figure A.3** – Locus des gènes caractéristiques des utilisations du sélénium chez les eubactéries et archaees séquencés en 2005.<sup>[A.1]</sup> Les gènes *selABC* se présentent sous la forme d'un opéron (jaune, bleu et vert) chez la plupart des organismes utilisant la sélénocystéine. L'indication de gène *selA* chez les trois archaees correspond en fait au gène de protéine Sela-like, SepSecS n'ayant pas encore été trouvée en 2005. (voir § 2.3.5.3, page 104)





A.4. Les séquences protéiques de *ecSclA* et de *mtSclA*



## A.5 Arbres phylogéniques des enzymes PLP-dépendantes

Les arbres phylogéniques des enzymes PLP-dépendantes, obtenus par différentes équipes,<sup>[A.4-A.7]</sup> proposent des placements des sélénocystéines synthèses incompatibles : cela montre bien la difficulté que représente la construction de la phylogénie de cette enzyme.

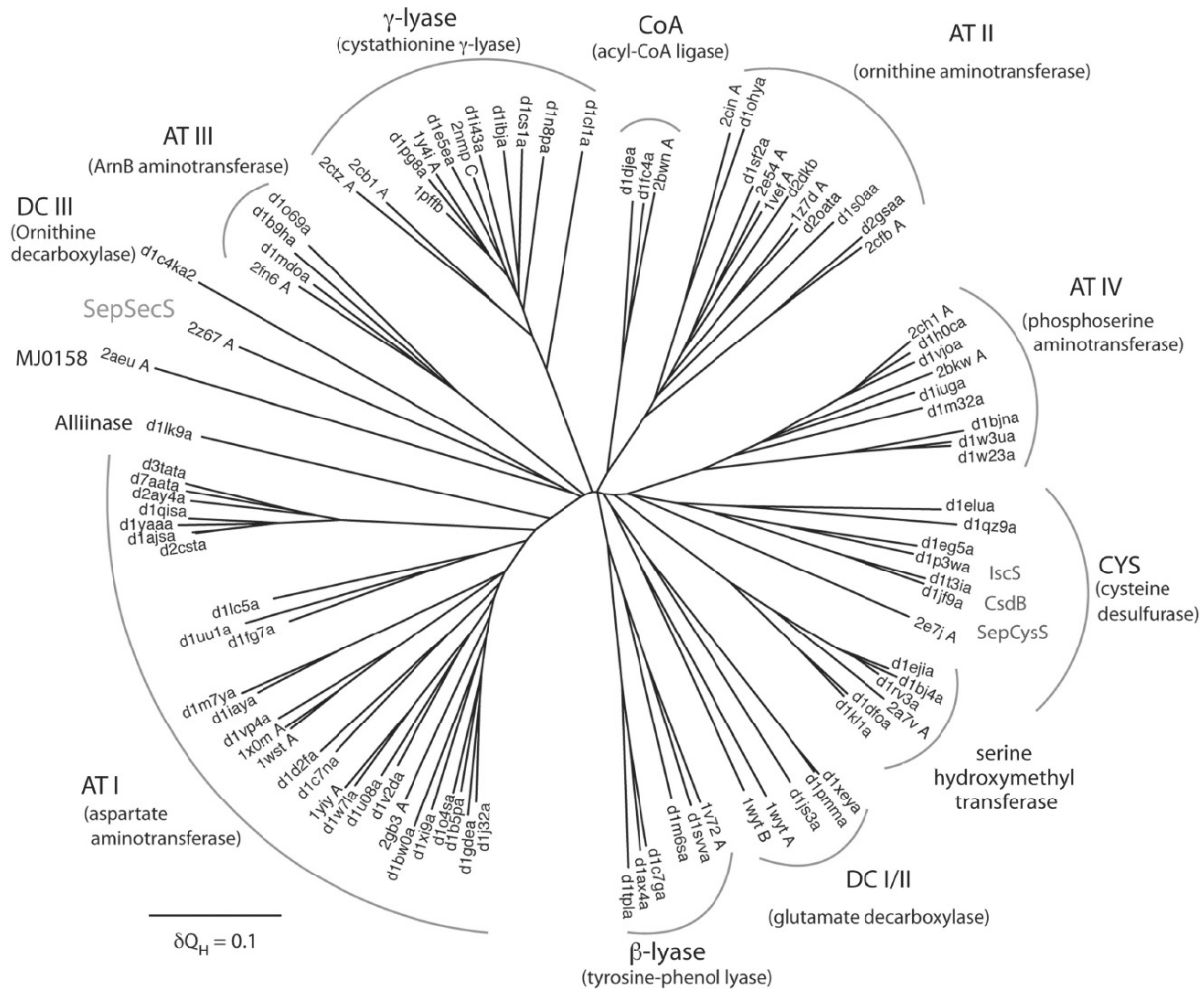


Figure A.5 – Arbre phylogénique proposé par Araiso *et al.*<sup>[A.7]</sup>

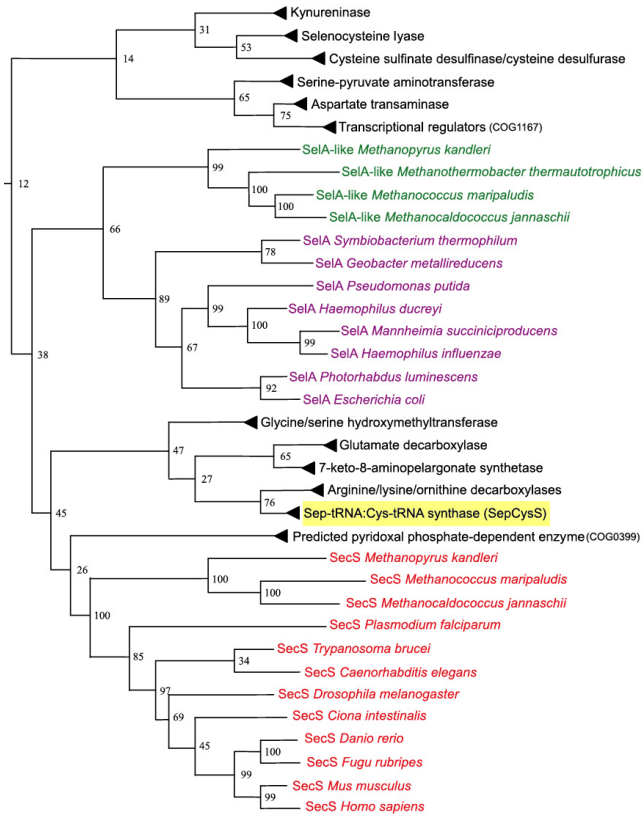


Figure A.6 – Arbre phylogénique proposé par Xu *et al.*<sup>[A.6]</sup>

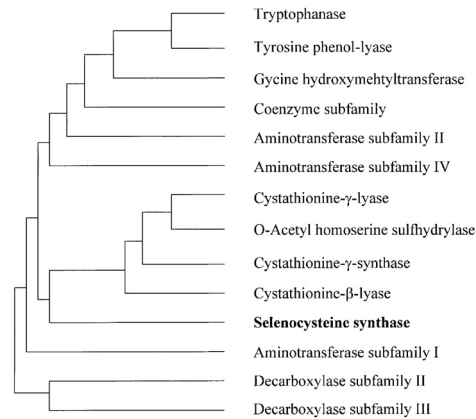
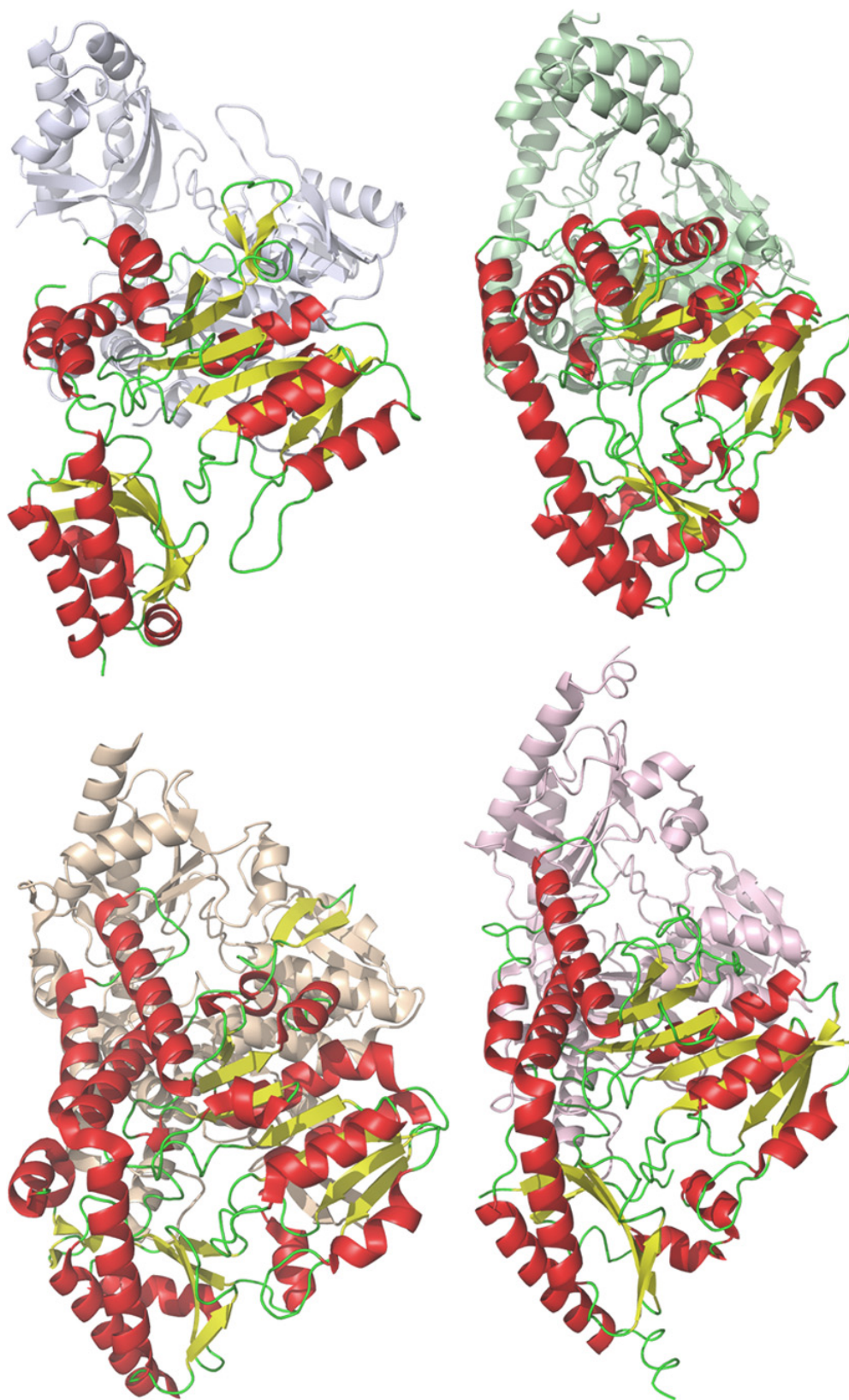


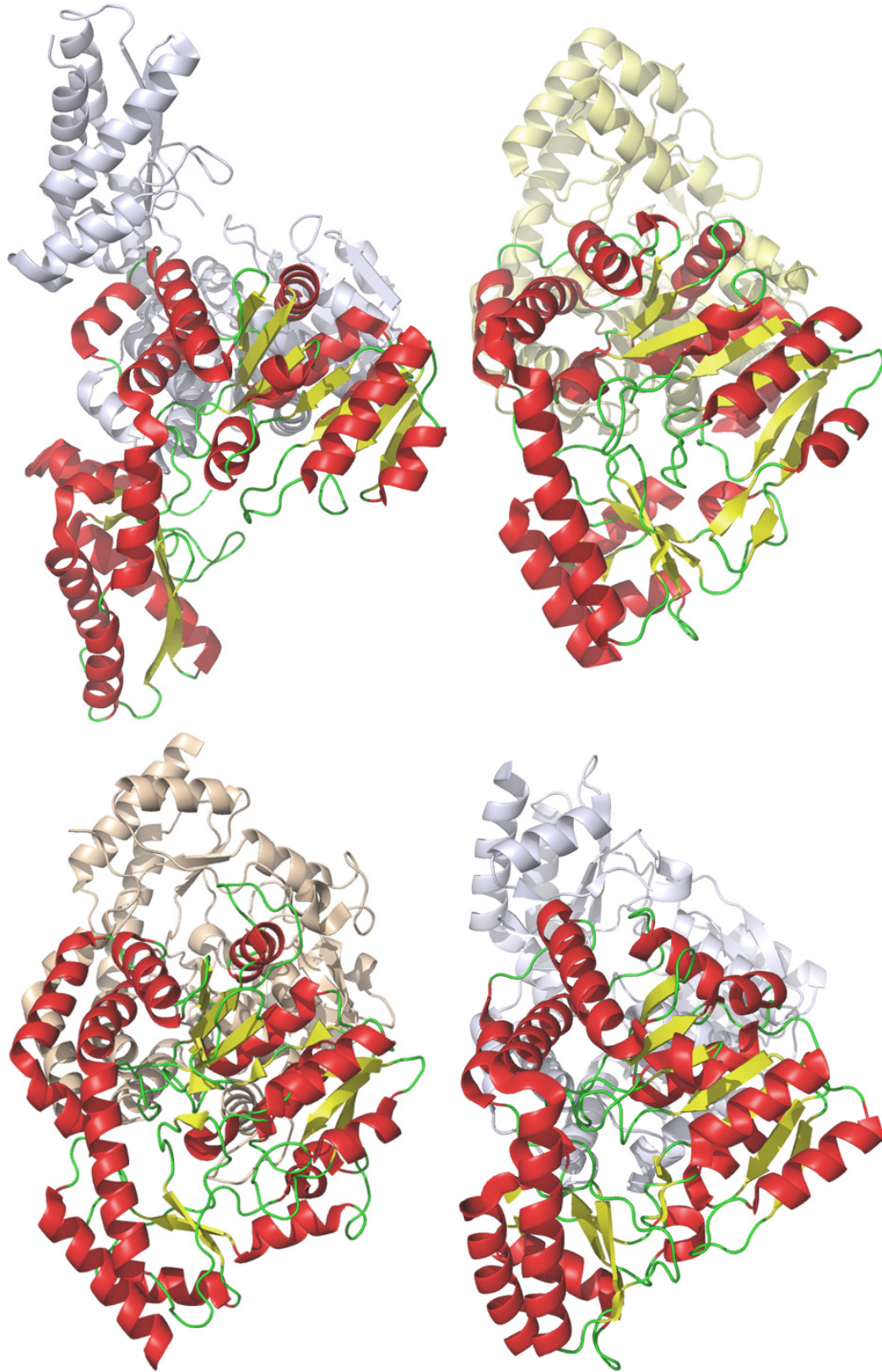
Figure A.7 – Arbre phylogénique proposé par Tormay *et al.*<sup>[A.4]</sup>

## A.6 Structures d'enzymes PLP-dépendantes

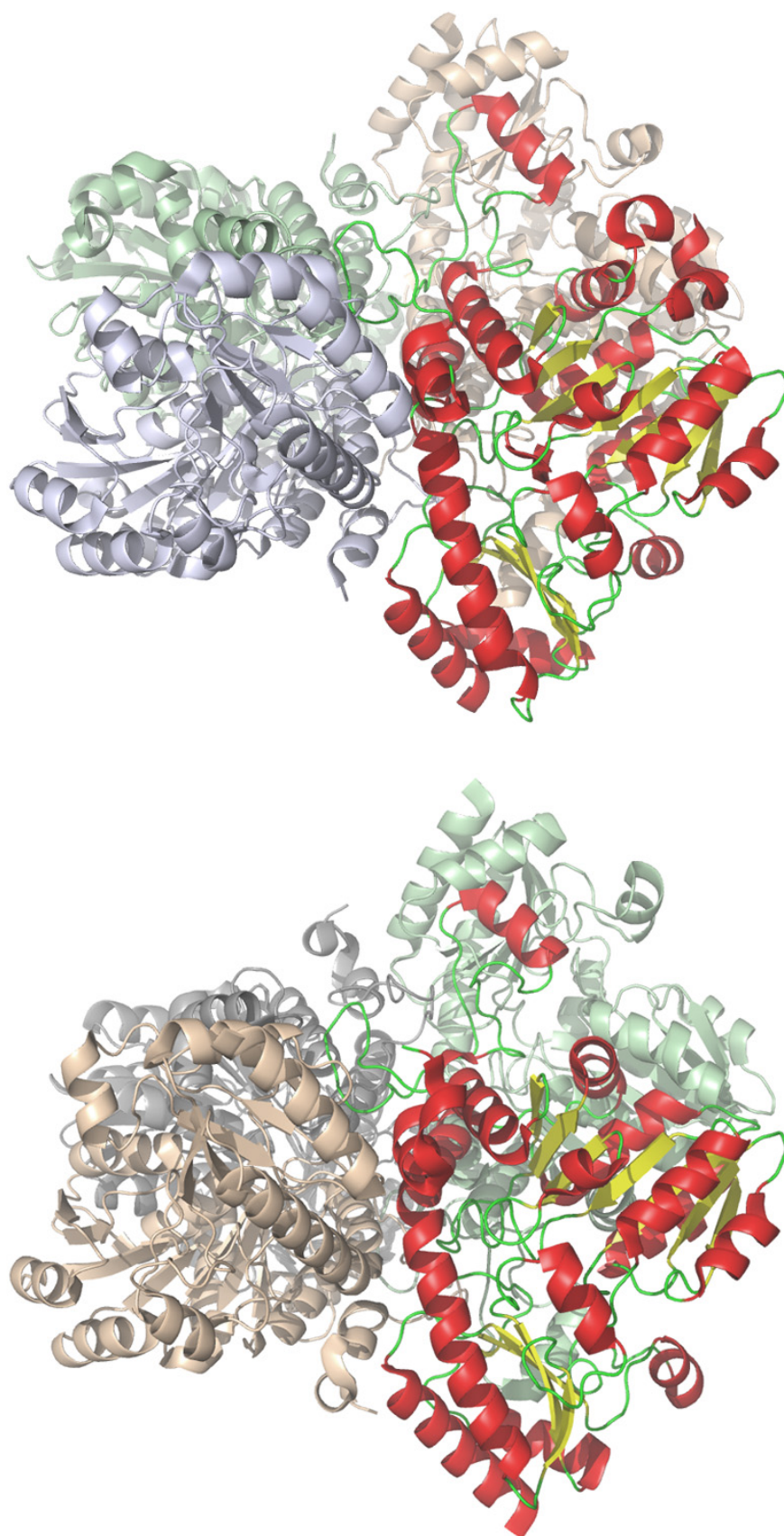
Sur les pages suivantes sont présentées les différentes enzymes PLP-dépendantes de «Fold Type I» rencontrées dans ce manuscrit, ou utilisées comme références lors des prédictions de structures secondaires de SelA.



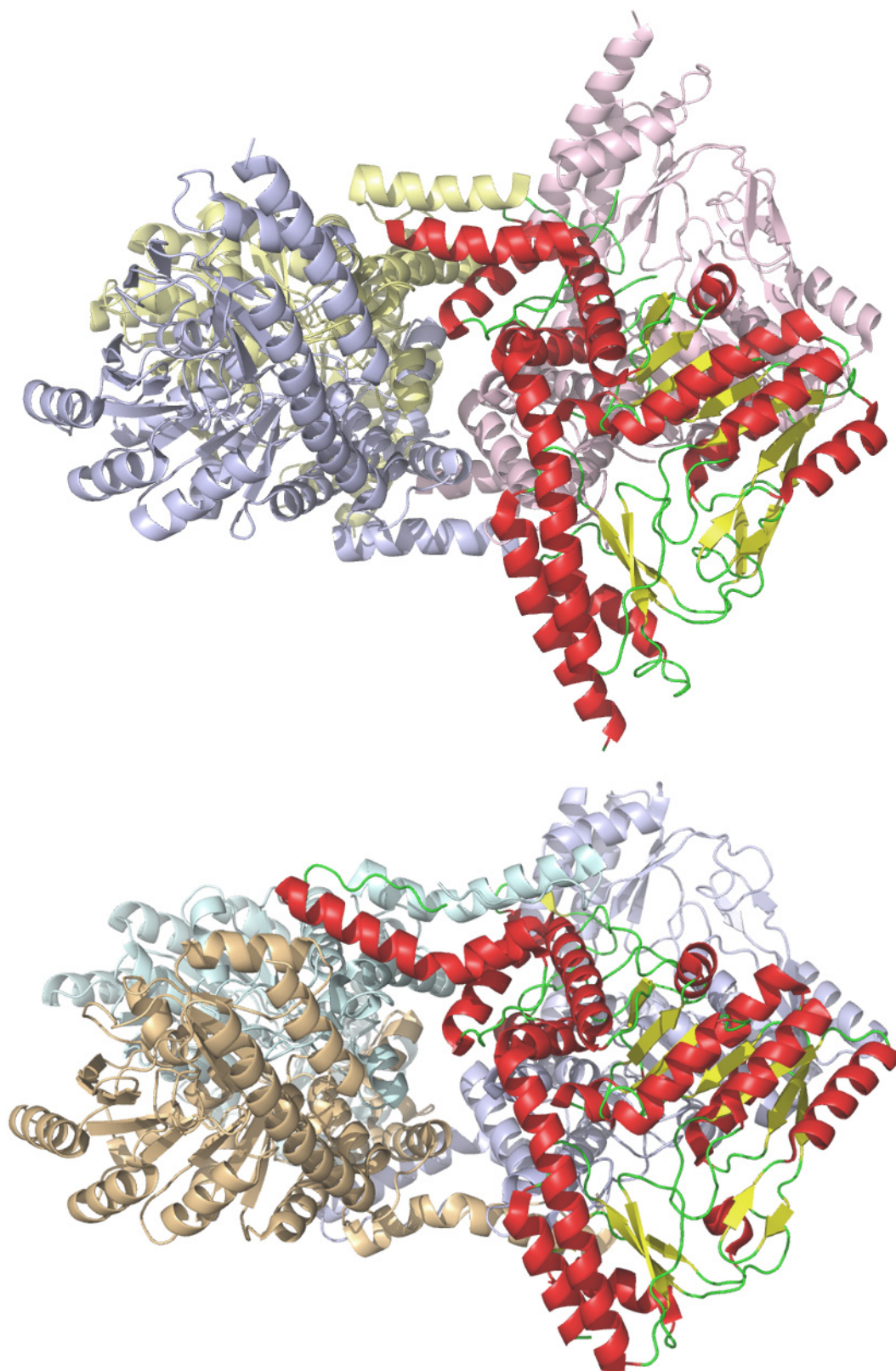
**Figure A.8** – (En haut à gauche) La NifS / CsdB d'*E. coli*, n°PDB : 1KMJ.  
(En bas à gauche) L'IscS d'*E. coli*, n°PDB : 1P3W.  
(En haut à droite) La SepCysS de *Archaeoglobus fulgidus*, n°PDB : 2E7J.  
(En bas à droite) L'aspartate aminotransférase de *Thermus thermophilus*, n°PDB : 1B5P.



**Figure A.9** – (En haut à gauche) La DesV de *Streptomyces venezuelae*, n°PDB : 2OGE.  
(En bas à gauche) L'alanine aminotransférase de *Thermus thermophilus*, n°PDB : 2YRR.  
(En haut à droite) La "Sela-like" MJ0158 de *M. jannaschii*, n°PDB : 2AEU.  
(En bas à droite) La glutamine aminotransférase de *Thermus thermophilus*, n°PDB : 1V2D.

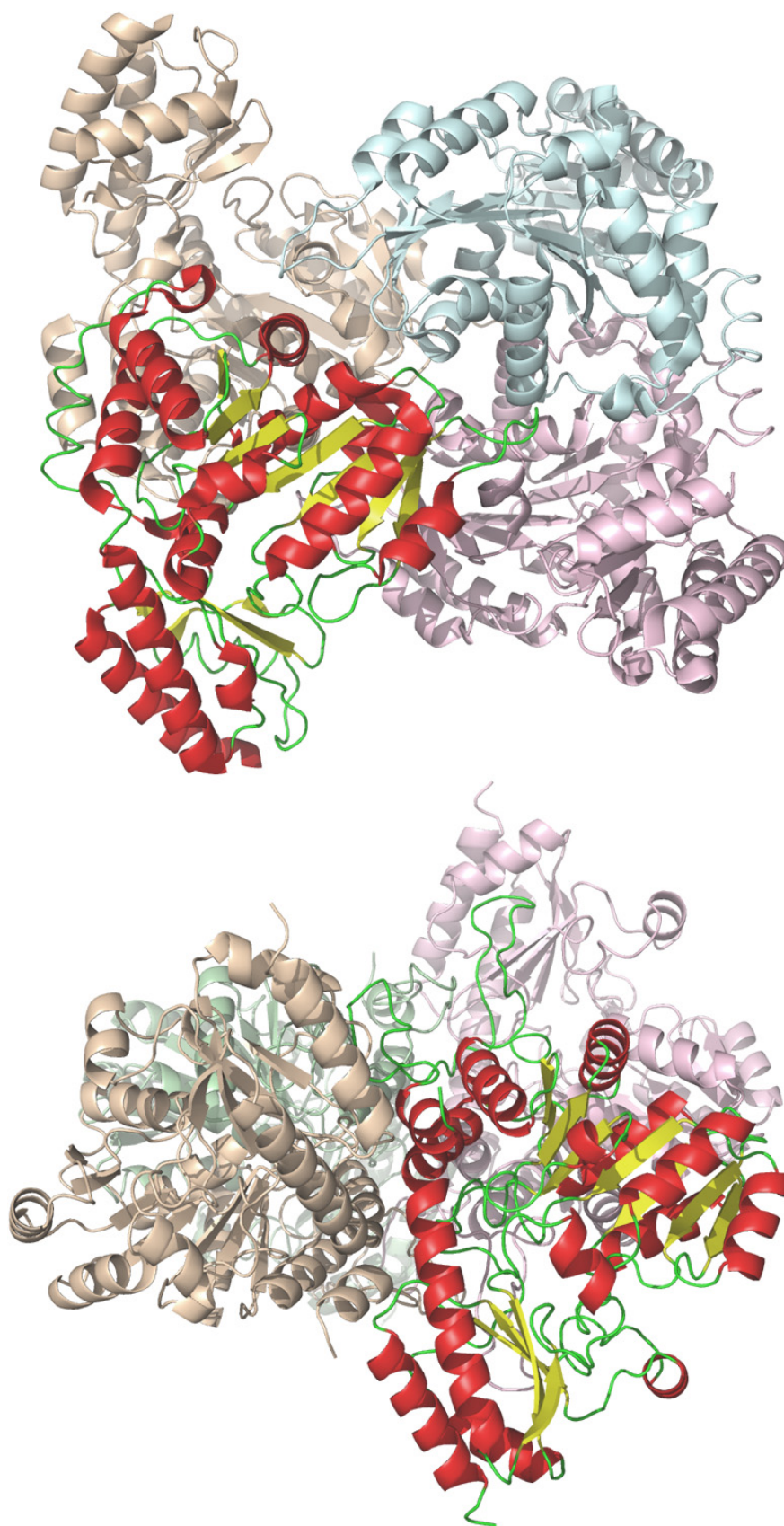


**Figure A.10** – (À gauche) La L-méthionine-lyase de *Citrobacter freundii*, n°PDB : 1Y4I, (À droite) La O-acétyl-L-homoserine-sulphydrylase de *Thermus thermophilus*, n°PDB : 2CB1.



**Figure A.11** – (À gauche) La SepSecS de *Methanococcus maripaludis*, n°PDB : 2Z67, (À droite) La SepSecS de *Mus musculus*, n°PDB : 3BC8.





**Figure A.12** – (Å gauche) La cystathionine  $\gamma$ -synthase (CGS) d'*Escherichia coli*, n°PDB : 1CS1, (Å droite) La thréonine aldolase de *Thermotoga maritima*, n°PDB : 1M6S. (Cette dernière possède une structure oligomérique particulière.)

## A.7 Structures des protéines décāmériques connues

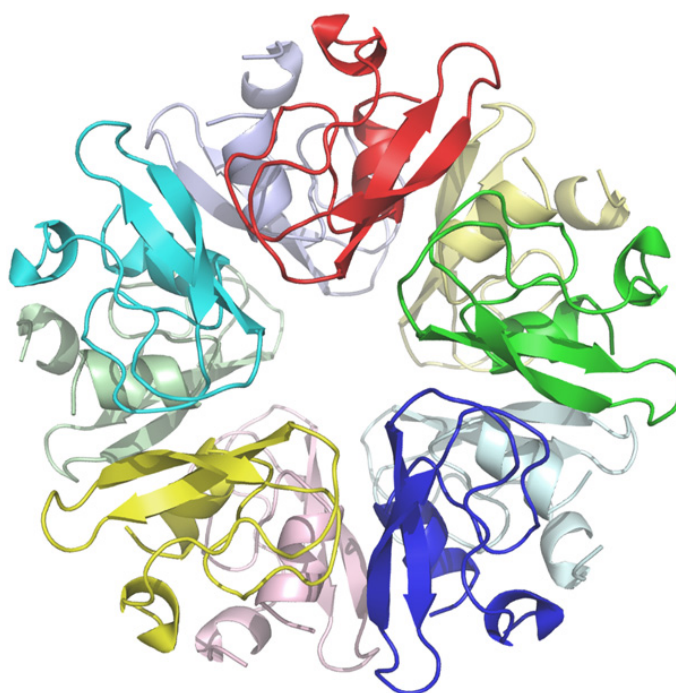
Voici les protéines décāmériques, exhibant une symétrie rotationnelle d'ordre 5, dont les structures ont été résolues (relatif au § 2.3.6, page 106) :

Protéine	Organisme	Masse du monomère	Identification PDB
Muconolactone isomérase	<i>Pseudomonas putida</i>	11 kDa	1MLI <sup>[A.8]</sup>
“Bovine pancreatic trypsin inhibitor”	<i>Bos taurus</i>	6,5 kDa	1BHC <sup>[A.9]</sup>
Cyanase	<i>Escherichia coli</i>	17 kDa	1DW9 <sup>[A.10]</sup>
2-Cys peroxyrédoxine	<i>Homo Sapiens</i>	22 kDa	1QMV <sup>[A.11]</sup>
Fructose-6-phosphate aldolase	<i>Escherichia coli</i>	25 kDa	1L6W <sup>[A.12]</sup>
GTP cyclohydrolase I	<i>Rattus norvegicus</i>	27 kDa	1IS7 <sup>[A.13]</sup>
Ketopantoate hydroxyméthyltransférase	<i>M. tuberculosis</i>	29 kDa	1OY0 <sup>[A.14]</sup>
Lectine galactose-spécifique	<i>Crotalus atrox</i>	15 kDa	1JZN <sup>[A.15]</sup>
Lumazine synthase (RibH2)	<i>Brucella abortus</i>	18 kDa	1XN1 <sup>[A.16]</sup>

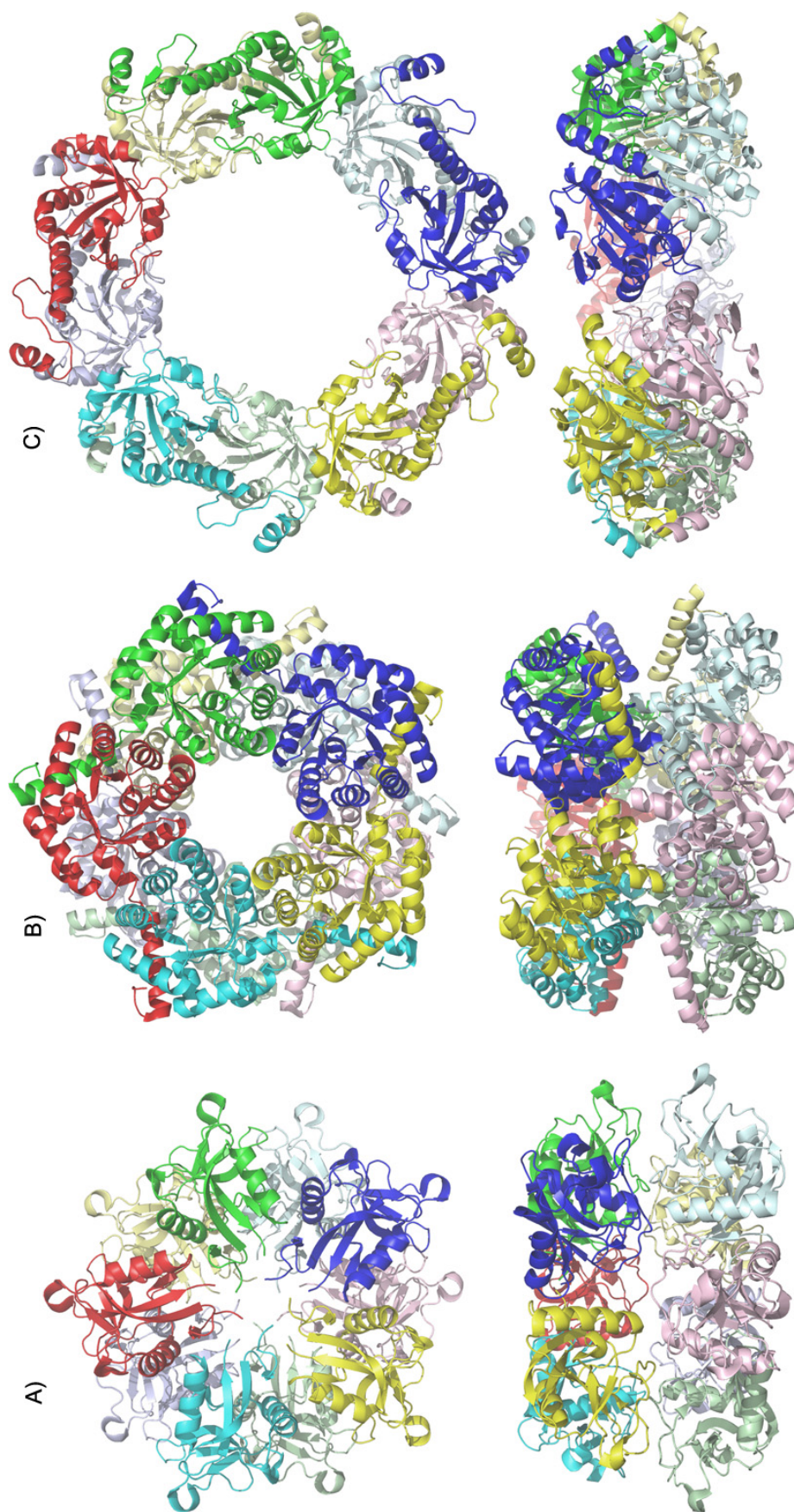
**Tableau A.1** – Protéines homo-décāmériques possédant une symétrie rotationnelle d'ordre 5.

Malgré le point commun d'être toutes des décāmères, ces protéines exhibent des variations d'organisations oligomériques. Les surfaces d'interaction entre deux unités d'un dimères et entre deux unités au sein d'un pentamères permettent de classer ces protéines décāmériques dans l'une des trois catégories suivantes : les pentamères de dimères (comme la 2-Cys peroxyrédoxine de *Homo Sapiens*, n° PDB : 1QMV, page suivante); les dimères de pentamères (soit la majorité d'entre elles); les protéines pour lesquelles la surface d'interaction est équivalente entre les deux unités d'un dimère et deux unités consécutives d'un pentamère.

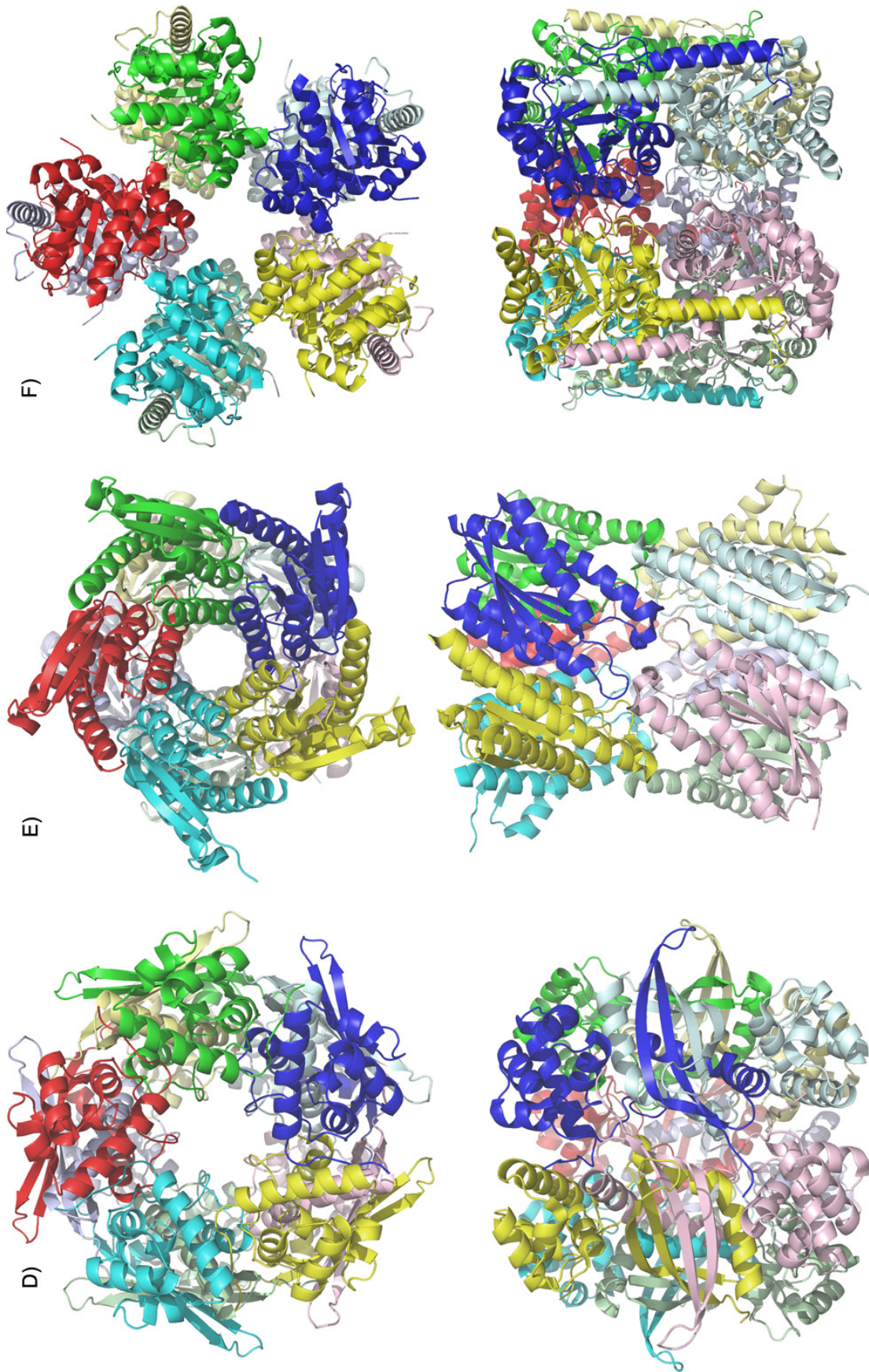
SelA peut être classée dans la catégorie des pentamères de dimères : la surface d'interaction au sein de ses dimères est la plus importante.



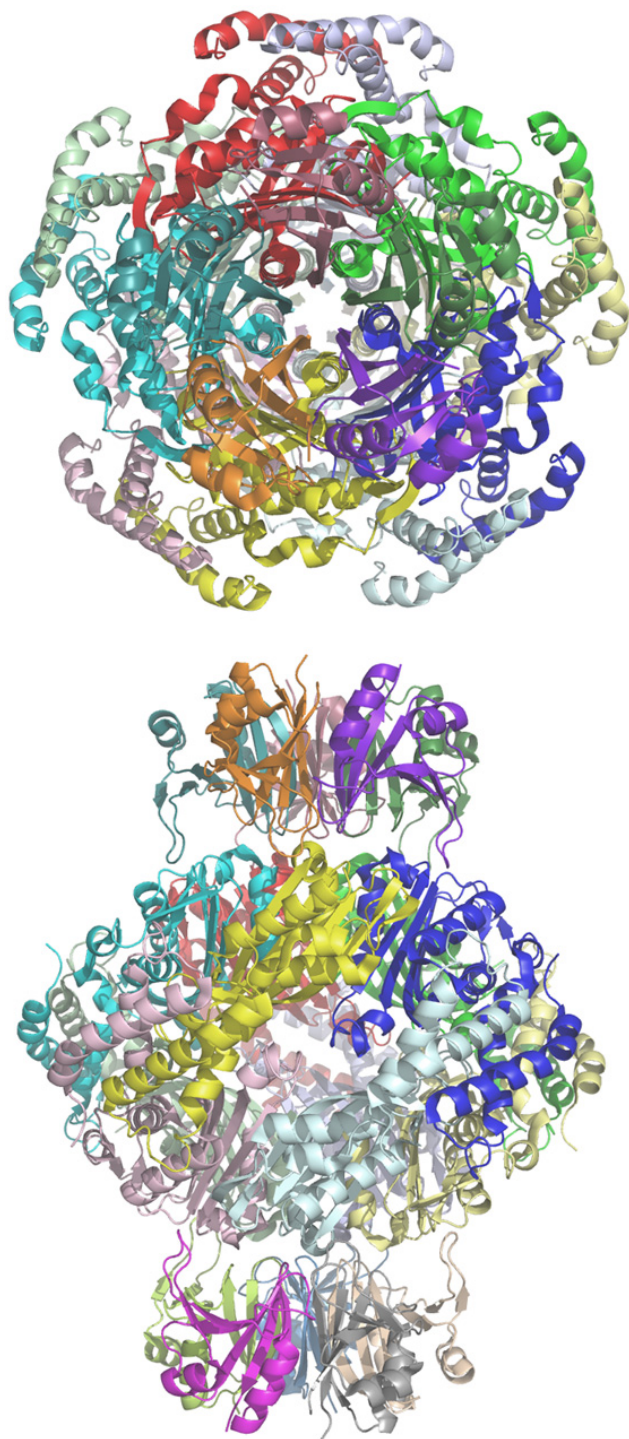
**Figure A.13** – Structure du “Bovine pancreatic trypsin inhibitor”, n° PDB : 1BHC.



**Figure A.14** – En A), la lectine galactose-spécifique de *Crotalus atrox*, n°PDB : 1JZN, En B), la fructose-6-phosphate aldolase d'*Escherichia coli*, n°PDB : 1L6W, En C), la 2-Cys peroxyrédoxine humaine, n°PDB : 1QMV.

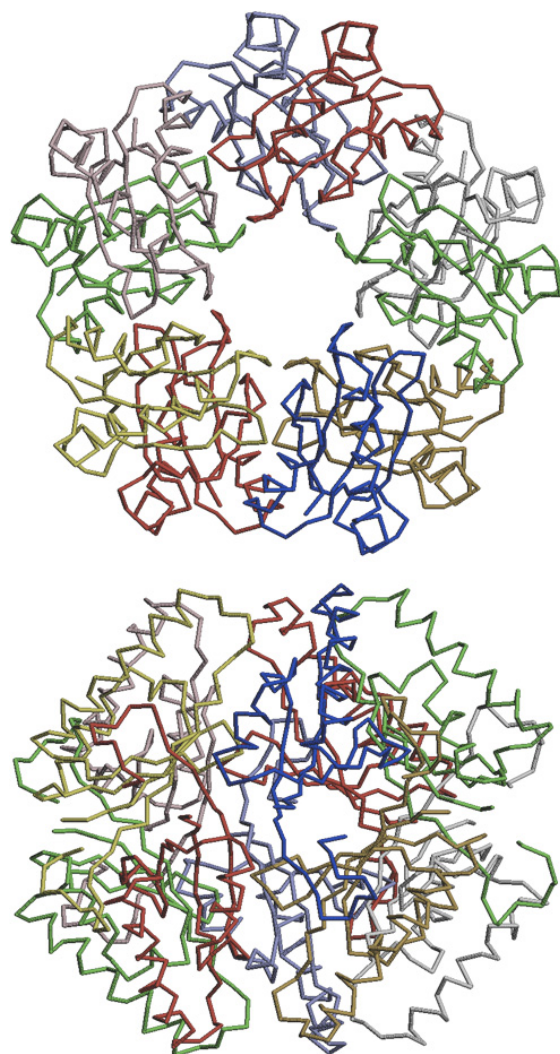


**Figure A.15** – En D), la cyanase d'*Escherichia coli*, n°PDB : 1DW9, En E), la lumazine synthase (RibH2) de *Brucella abortus*, n°PDB : 1XN1, En F), la ketopantoate hydroxyméthyltransférase de *Mycobacterium tuberculosis*, n°PDB : 1OY0.



**Figure A.16** – Complexe de la GTP cyclase I de *Rattus norvegicus*, avec sa protéine régulatrice GFRP, n°PDB : 1IS7.

La GTP cyclase I (figure A.16) est une exception dans le paysage de ces protéines décimériques : ce décimère est ici en complexe avec deux pentamères de sa protéine régulatrice GFRP, qui stimule son activité vers la synthèse de tétrahydrobioptérine en présence de phénylalanine.



**Figure A.17** – La muconolactone isomérase de *Pseudomonas putida*, n°PDB : 1MLI. La résolution de la densité électronique obtenue par diffraction aux rayons X ne permet de placer que les C $\alpha$  de la protéine.

# Annexe B

## Matériels et Méthodes

### B.1 Fiches techniques

#### B.1.1 Plasmide utilisé pour le clonage de *selC*

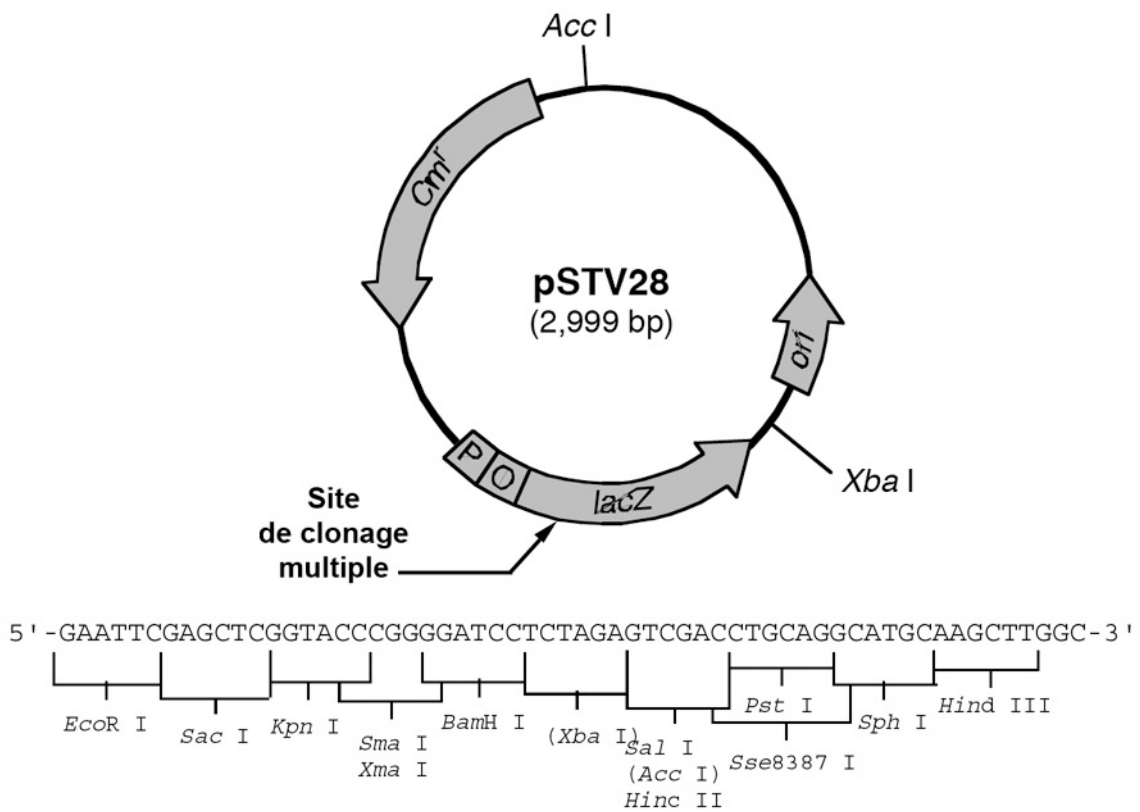


Figure B.1 – Carte du plasmide pSTV28. (Produit par TaKaRa Biotechnology®).



## B.2 Protocoles expérimentaux

Ma mutation de la chimie à la biologie m'a amené à découvrir, à utiliser puis à maîtriser un certain nombre de protocoles de biologie moléculaire. Certains pourront trouver que les fiches de protocole suivantes sont triviales ou concernent des protocoles basiques, et ils auront raison. Elles m'ont pourtant été personnellement fort utiles, dans un domaine où j'ai eu tout à apprendre, dans un contexte nécessitant une reproductibilité et une rigueur constante.

### B.2.1 PCR de multiplication de la matrice de transcription

Volume type = 1mL (multiple de 100 $\mu$ L) Valable uniquement pour *selC* d'*E. coli*.

**Tableau B.1** – Solutions nécessaires et mélange réactionnel de PCR.

Tampon Taq 10x (Thermopol Taq buffer <i>ou</i> Standard Taq buffer)	10 $\mu$ L
dNTPs 2mM (Total)	10 $\mu$ L
MgCl <sub>2</sub> 25mM	6 $\mu$ L
Matrice ADN à multiplier	1ng/mL final
Primer 100 $\mu$ M	0,4 $\mu$ L x2
Taq ADNpolymerase (sortir au dernier moment du -20°C)	1 $\mu$ L
H2O SMQ	qsp 100 $\mu$ L

*Nota* : ThermoPol Taq buffer 10x = (200mM TrisHCl pH 8,8)+(100mM KCl)+(100mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>)+(20mM MgSO<sub>4</sub>)+(1% Triton 100x)

Standard Taq buffer 10x = (100mM TrisHCl pH 8,3)+(500mM KCl)+(15mM MgCl<sub>2</sub>)+(1% Triton 100x)

#### Mode opératoire :

- Mélanger le tout dans un tube.
- Disposer des volumes de 100 $\mu$ L dans des tubes PCR.
- Utiliser le programme adapté TAQ (durée environ 1H20) :

- 1 : 2min à 92°C
- 2 : 20sec à 92°C
- 3 : 30sec à 55°C
- 4 : 10sec à 72°C
- 5 : GOTO2 REP25
- 6 : 4min à 72°C
- 7 : 10°C until STOP



- Rassembler dans un ou plusieurs tube(s) Ependorf (remplissage au tier).
- Extraire au mélange phénol/chloroforme/isoamyl alcohol (25:24:1) saturé en Tris 10mM pH 8,0 et 1mM d'EDTA, puis au chloroforme (1 Volume à chaque fois).
- Précipiter dans de l'EtOH (2,5 Volumes) + Tris 10mM (final) à pH 8,0 : laisser 4H minimum à -20°C.
- Centrifuger les tubes Ependorf à 12000 RPM à 4°C (Centrifugeuse de paillasse), 20 minutes.
- Après avoir retiré le surnageant, lyophiliser au SpeedVac.
- Solubiliser dans du Tris 10mM pH 8,0 (au volume de PCR).
- *Estimer* la quantité obtenue grâce à l'absorbance à 260nm (en diluant d=10) :  
 $A=1 \rightarrow C=50 \text{ ng}/\mu\text{L}$
- Vérifier le résultat sur un gel d'agarose à 2%, en comparant l'intensité de la bande obtenue, à différente dilution de l'échantillon, avec les bandes du marqueur.

### B.2.2 Transcription de *selC* en ARN<sub>t</sub><sup>Sec</sup>

Volume type = 10mL (Volume multiple de 1mL)

Préférer la deuxième méthode (Tableau B.3, page ci-contre).

**Tableau B.2** – Méthode de Nicolas - Mélange de solutions nécessaires pour la transcription.

Tris pH 8,0	2M	20 $\mu\text{L}$
MgCl <sub>2</sub>	1M	10 $\mu\text{L}$
DTT	1M	5 $\mu\text{L}$
Spermidine	500mM	0,2 $\mu\text{L}$
NTPs (25mM chacun)		15 $\mu\text{L}$
GMP	100mM	36 $\mu\text{L}$
Matrice ADN de <i>selC</i>		(concentration finale en ADN de 12 $\mu\text{g}/\text{ml}$ )
T7polymerase		10 $\mu\text{L}$
	(sortir au dernier moment du -20°C)	
H2O SMQ		qsp 1mL

#### Mode opératoire :

- Mélanger le tout, dans l'ordre, dans un tube approprié.
- Disposer dans le bain-marie à 37°C pendant 4H.
- Après la réaction, ajouter 0,1 Volume d'EDTA à 0,5M, mettre sur glace.
- Prélever 10 $\mu\text{L}$  pour faire une vérification sur Gel de polyacrylamide.
- Extraire au phénol puis au chloroforme (1 Volume à chaque fois).
- Précipiter dans de l'EtOH (2,5 Volumes) + AcONa 3M pH 5,0 (0,1 Volume) dans un tube Corning : laisser au minimum 4H à -20°C (au mieux toute la nuit).
- Centrifuger le tube Corning à 12000 RPM à 4°C (Rotor JA14) (= 22000g), 20 minutes.
- Lyophiliser, puis resolubiliser dans 150 $\mu\text{L}$  de AcONa 50mM pH 5,0.

**Tableau B.3** – Méthode de Dominique - Mélange de solutions nécessaires pour la transcription.

HEPES pH 7,6	1M	20 $\mu$ L
MgCl <sub>2</sub>	1M	20 $\mu$ L
DTT	1M	5 $\mu$ L
Spermidine	500mM	2 $\mu$ L
NTPs (25mM chacun)		40 $\mu$ L
GMP	100mM	100 $\mu$ L
Matrice ADN de <i>selC</i>		(concentration finale en ADN de 12 $\mu$ g/ml)
<b>Pyrophosphatase</b>		10 unités
T7polymerase (sortir au dernier moment du -20°C)		4 $\mu$ L
H2O SMQ		qsp 1mL

- Mesurer la quantité obtenue grâce à l'absorbance à 260nm (en diluant d=40) :

$$C=A/\epsilon \rightarrow m=AxVxM/\epsilon$$

Pour l'ARN<sub>t</sub><sup>Sec</sup> d'*E. coli* :  $\epsilon=896100$  et  $M=30568\text{Da}$ .

### B.2.3 Charge en sérine de l'ARN<sub>t</sub><sup>Sec</sup>

#### Charge chaude à la sérine <sup>14</sup>C - Étude cinétique

Volume d'un échantillon de charge : 50 $\mu$ L.

Réaction type = 3x50 $\mu$ L + 20 $\mu$ L de Témoin

**Tableau B.4** – Solutions nécessaires et mélange de charge chaude.

Concentration finale d'ARN <sub>t</sub> :	0,4 $\mu$ M	0,7 $\mu$ M	1,0 $\mu$ M	T-
Pure System Buffer 12% (8,33x)	6 $\mu$ L	6 $\mu$ L	6 $\mu$ L	2,4 $\mu$ L
ARN <sub>t</sub> <sup>Sec</sup> à 4 $\mu$ M (Snapcoolé)	5 $\mu$ L	8,8 $\mu$ L	12,5 $\mu$ L	/
ATP 100mM	1 $\mu$ L	1 $\mu$ L	1 $\mu$ L	0,4 $\mu$ L
Sérine 14C 637 $\mu$ M	5 $\mu$ L	5 $\mu$ L	5 $\mu$ L	2 $\mu$ L
<b>Pyrophosphatase</b>	0,25 unités	0,25 unités	0,25 unités	/
SerRS (à 5mg/mL) (sortir au dernier moment du -20°C)	2,8 $\mu$ L	2,8 $\mu$ L	2,8 $\mu$ L	1,2 $\mu$ L
H <sub>2</sub> O SMQ	qsp 50 $\mu$ L	qsp 50 $\mu$ L	qsp 50 $\mu$ L	qsp 20 $\mu$ L
<i>Pour un échantillonnage de 15<math>\mu</math>L</i>	1600 CPM	2800 CPM	4000 CPM	0 CPM

#### Mode opératoire :

- Snapcooler l'ARN<sub>t</sub> à 4 $\mu$ M : 1min à 95°C et 5min sur glace.

- Mélanger le tout (sauf SerRS), dans l'ordre, dans un tube Ependorf 1,5mL.

- Ajouter la SerRS en déclanchant le chronomètre et disposer dans le bain-marie à 37°C.
- À 10, 15 et 20min, centrifuger, prélever 15µL, déposer sur des petits carrés de papier Wattman (numérotés et épinglés sur support) et plonger les directement dans du TCA à 5% (l'ARN précipite et reste sur le papier, la sérine 14C résiduelle est lavée).
- Laver 2x dans du TCA 5% 30min, puis dans de l'EtOH 30min.
- Sécher les papiers, les placer dans des cuves contenant du liquide à scintillation, Compter !

### Charge froide - Production

(à la concentration d'ARN<sub>t</sub><sup>Sec</sup> déterminée par la manip de charge chaude)

Volume type = 10mL (Volume multiple de 1mL)

**Tableau B.5** – Solutions nécessaires et mélange de charge froide.

Concentration finale d'ARNt :	0,4µM	0,7µM	1,0µM
Pure System Buffer 12% (8,33x)	120µL	120µL	120µL
ARN <sub>t</sub> <sup>Sec</sup> à 4µM (Snapcoolé)	100µL	175µL	250µL
ATP 100mM	20µL	20µL	20µL
Sérine 1mM	64µL	64µL	64µL
<b>Pyrophosphatase</b>	5 unités	5 unités	5 unités
SerRS (à 5mg/mL)	56µL	56µL	56µL
(sortir au dernier moment du -20°C)			
H <sub>2</sub> O SMQ	1mL	1mL	1mL

### Mode opératoire :

- Snapcooler l'ARN<sub>t</sub><sup>Sec</sup> à 4µM : 1min à 95°C et 5min sur glace.
- Mélanger le tout, dans l'ordre, dans un tube de taille appropriée.
- Disposer dans le bain-marie à 37°C pendant 15min (ou moins selon l'étude cinétique).
- Après la réaction, ajouter 0,1 Volume d'AcONa 3M à pH 5,5 et mettre sur glace.
- Extraire au Phénol puis au Chloroforme (1 Volume à chaque fois).
- Précipiter dans de l'EtOH (2,5 Volumes) + AcONa 3M pH 5,5 (0,02 Volume) dans un tube Corning : laisser 4H max, à -20°C.
- Centrifuger le tube Corning à 12 000 RPM à 4°C (Rotor JA14) (= 22000G).
- Laver avec EtOH 70%, puis re-centrifuger.
- Lyophiliser, puis resolubiliser dans 100µL de AcONa 50mM pH 5,0.

### Nota Bene :

La quantité d'enzyme utilisée prend en compte la cinétique de charge du nouveau lot de SerRS (SG).

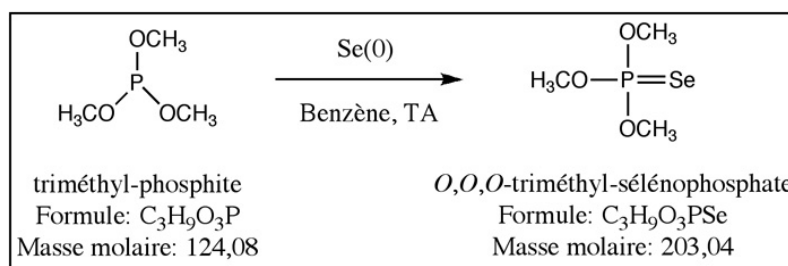
# Annexe C

## La chimie du sélénium

### Synthèse chimique du sélénophosphate $\text{SePO}_3^{3-}$

Cette synthèse s'effectue en trois étapes : obtention de l'intermédiaire tris(triméthylsilyl)-sélénophosphate  $(\text{TMSO})_3\text{PSe}$  en deux étapes présentées ci-dessous, et hydrolyse de cet intermédiaire en solution dans du chloroforme, par un transfert de phase dans un tampon aqueux adéquate (protocoles selon Glass *et al.*<sup>[A.17,A.18]</sup> et Bhardwaj *et al.*<sup>[A.19]</sup>).

### Synthèse du triméthyl-sélénophosphate $(\text{CH}_3\text{O})_3\text{PSe}$



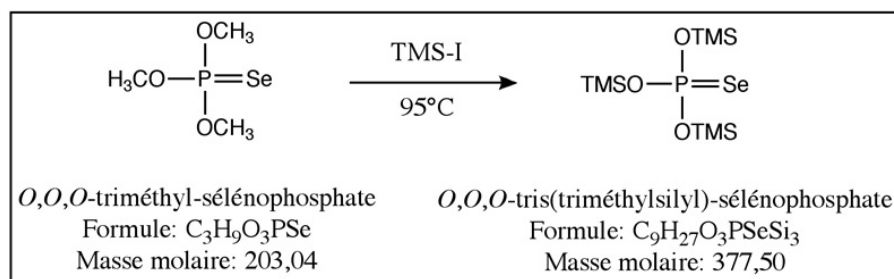
À une suspension de poudre de sélénium élémentaire (2,0 g ; 25,3 mmol) dans du benzène anhydre (5,0 mL) sous agitation magnétique, est ajouté la triméthyl-phosphite (2,50 mL ;  $d=1,052$  ; 2,63 g ; 21,2 mmol), à température ambiante, sous azote. Après 24h, le milieu réactionnel est filtré sur celite. Le benzène est distillé sous pression atmosphérique, et le résidu distillé sous vide pour aboutir au *O,O,O*-triméthyl-sélénophosphate (2,51 g ; 12,4 mmol) avec un rendement de 58,5%.

*O,O,O*-triméthyl-sélénophosphate  $(\text{CH}_3\text{O})_3\text{PSe}$  :

RMN  $^1\text{H}$  ( $\text{CDCl}_3$ , 300,17 MHz) :  $\delta = 3,75$  ppm (d,  $^3\text{J}_{\text{H-C-O-P}} = 13,3$  Hz, 9 H)

RMN  $^{13}\text{C}$  ( $\text{CDCl}_3$ , 75,48 MHz) :  $\delta = 55,08$  ppm (s, 3 C)

RMN  $^{31}\text{P}$  ( $\text{CDCl}_3$ , 121,50 MHz) :  $\delta = 78,04$  ppm (t,  $^1\text{J}_{\text{P-}^{77}\text{Se}} = 464$  Hz, 1 P)

Synthèse du tris(triméthylsilyl)-sélénophosphate (TMSO)<sub>3</sub>PSe

Au triméthyl-sélénophosphate (1,59 g ; 7,83 mmol) sous agitation magnétique, est ajouté l'iodure de triméthylsilyl (3,40 mL ; d=1,471 ; 5,00 g ; 25,0 mmol), à température ambiante, sous azote. Le mélange agité est chauffé à 95°C pendant 24h, puis séparé par distillation fractionnée tout d'abord sous pression atmosphérique, puis sous vide, pour aboutir au tris(triméthylsilyl)-sélénophosphate (2,76 g ; 7,31 mmol) avec un rendement de 93,4%. La réaction conduit à la formation d'iodométhane MeI, sous-produit qui peut être distillé hors du milieu réactionnel au cours de la réaction (T°eb=43°C).

*O,O,O*-tris(triméthylsilyl)-sélénophosphate (TMSO)<sub>3</sub>PSe :

RMN <sup>1</sup>H (CDCl<sub>3</sub>, 300,17 MHz) : δ = 0,28 ppm (s, 27 H)

RMN <sup>13</sup>C (CDCl<sub>3</sub>, 75,48 MHz) : δ = 1,01 ppm (s, 9 C)

RMN <sup>31</sup>P (CDCl<sub>3</sub>, 121,50 MHz) : δ = 21,45 ppm (t, <sup>1</sup>J<sub>P-<sup>77</sup>Se</sub> = 462 Hz, 1 P)

## Hydrolyse du tris(triméthylsilyl)-sélénophosphate

Pour obtenir une solution de sélénophosphate prête à l'emploi, sous atmosphère inerte, mélanger vigoureusement un volume de solution de tris(triméthylsilyl)-sélénophosphate (TMSO)<sub>3</sub>PSe dans le chloroforme à 120mM, avec un volume identique de tampon dégazé Tricine-KOH à 100mM, contenant 60mM de MgCl<sub>2</sub> et 20mM de DTT, à pH 7,2.

Le pH de la solution obtenue peut être corrigé vers la valeur désirée par ajout de solution concentrée d'hydroxyde de potassium KOH ou d'acide chlorhydrique HCl.

# Références bibliographiques des Annexes

- [1] ROMERO H. ; ZHANG Y. ; GLADYSHEV V. N. ; SALINAS G. : Evolution of selenium utilization traits. *Genome biology*, 6(8) : R66, 2005.
- [2] SAWERS G. ; HEIDER J. ; ZEHELEIN E. ; BÖCK A. : Expression and operon structure of the sel genes of *Escherichia coli* and identification of a third selenium-containing formate dehydrogenase isoenzyme. *Journal of bacteriology*, 173(16) : p4983–4993, Aug 1991.
- [3] ZHANG Y. ; ROMERO H. ; SALINAS G. ; GLADYSHEV V. N. : Dynamic evolution of selenocysteine utilization in bacteria : a balance between selenoprotein loss and evolution of selenocysteine from redox active cysteine residues. *Genome biology*, 7(10) : R94, 2006.
- [4] TORMAY P. ; WILTING R. ; LOTTSPPEICH F. ; MEHTA P. K. ; CHRISTEN P. ; BÖCK A. : Bacterial selenocysteine synthase—structural and functional properties. *European journal of biochemistry / FEBS*, 254(3) : p655–661, Jun 1998.
- [5] MEHTA P. K. ; CHRISTEN P. : The molecular evolution of pyridoxal-5'-phosphate-dependent enzymes. *Advances in enzymology and related areas of molecular biology*, 74 : p129–184, 2000.
- [6] XU X.-M. ; CARLSON B. A. ; MIX H. ; ZHANG Y. ; SAIRA K. ; GLASS R. S. ; BERRY M. J. ; GLADYSHEV V. N. ; HATFIELD D. L. : Biosynthesis of selenocysteine on its tRNA in eukaryotes. *PLoS biology*, 5(1) : e4, Jan 2007.
- [7] ARAISO Y. ; PALIOURA S. ; ISHITANI R. ; SHERRER R. L. ; O'DONOGHUE P. ; YUAN J. ; OSHIKANE H. ; DOMAE N. ; DEFRANCO J. ; SÖLL D. ; NUREKI O. : Structural insights into RNA-dependent eukaryal and archaeal selenocysteine formation. *Nucleic acids research*, 36(4) : p1187–1199, Mar 2008.
- [8] KATTI S. K. ; KATZ B. A. ; WYCKOFF H. W. : Crystal structure of muconolactone isomerase at 3.3 Å resolution. *Journal of molecular biology*, 205(3) : p557–571, Feb 1989.
- [9] HAMIAUX C. ; PRANGÉ T. ; RIÈS-KAUTT M. ; DUCRUIX A. ; LAFONT S. ; ASTIER J. P. ; VEESLER S. : The decameric structure of bovine pancreatic trypsin inhibitor (BPTI) crystallized from thiocyanate at 2.7 Å resolution. *Acta crystallographica. Section D, Biological crystallography*, 55(Pt 1) : p103–113, Jan 1999.
- [10] WALSH M. A. ; OTWINOWSKI Z. ; PERRAKIS A. ; ANDERSON P. M. ; JOACHIMIAK A. : Structure of cyanase reveals that a novel dimeric and decameric arrangement of subunits is required for formation of the enzyme active site. *Structure*, 8(5) : p505–514, May 2000.
- [11] SCHRÖDER E. ; LITTLECHILD J. A. ; LEBEDEV A. A. ; ERRINGTON N. ; VAGIN A. A. ; ISUPOV M. N. : Crystal structure of decameric 2-Cys peroxiredoxin from human erythrocytes at 1.7 Å resolution. *Structure*, 8(6) : p605–615, Jun 2000.
- [12] THORELL S. ; SCHÜRMMANN M. ; SPRENGER G. A. ; SCHNEIDER G. : Crystal structure of decameric fructose-6-phosphate aldolase from *Escherichia coli* reveals inter-subunit helix swapping as a structural basis for assembly differences in the transaldolase family. *Journal of molecular biology*, 319(1) : p161–171, May 2002.

- [13] MAITA N. ; OKADA K. ; HATAKEYAMA K. ; HAKOSHIMA T. : Crystal structure of the stimulatory complex of GTP cyclohydrolase I and its feedback regulatory protein GFRP. *Proceedings of the National Academy of Sciences of the United States of America*, 99(3) : p1212–1217, Feb 2002.
- [14] CHAUDHURI B. N. ; SAWAYA M. R. ; KIM C. Y. ; WALDO G. S. ; PARK M. S. ; TERWILLIGER T. C. ; YEATES T. O. : The crystal structure of the first enzyme in the pantothenate biosynthetic pathway, ketopantoate hydroxymethyltransferase, from *M. tuberculosis*. *Structure*, 11(7) : p753–764, Jul 2003.
- [15] WALKER J. R. ; NAGAR B. ; YOUNG N. M. ; HIRAMA T. ; RINI J. M. : X-ray crystal structure of a galactose-specific C-type lectin possessing a novel decameric quaternary structure. *Biochemistry*, 43(13) : p3783–3792, Apr 2004.
- [16] KLINKE S. ; ZYLBERMAN V. ; VEGA D. R. ; GUIMARÃES B. G. ; BRADEN B. C. ; GOLDBAUM F. A. : Crystallographic studies on decameric *Brucella* spp. Lumazine synthase : a novel quaternary arrangement evolved for a new function ? *Journal of molecular biology*, 353(1) : p124–137, Oct 2005.
- [17] GLASS R. S. ; SINGH W. P. ; JUNG W. ; VERES Z. ; SCHOLZ T. D. ; STADTMAN T. C. : Mono-selenophosphate : synthesis, characterization, and identity with the prokaryotic biological selenium donor, compound SePX. *Biochemistry*, 32(47) : p12555–12559, Nov 1993.
- [18] GLASS R. S. ; STADTMAN T. C. : Selenophosphate. *Methods in enzymology*, 252 : p309–315, 1995.
- [19] BHARDWAJ R. K. ; DAVIDSON R. S. : Oxidative desulphurisation and deselenation at pentavalent phosphorous by photogenerated peroxidic species. *Tetrahedron*, 43(19) : p4473–4479, 1987.

# Index

- Applications médicales, 24  
Apports journaliers et déficience, 21  
ARN<sub>t</sub> sélénés, 21, 30, 76  
ARN<sub>t</sub><sup>Sec</sup>, 41, 81, 102  
Asparagine (Asn) et Glutamine (Gln), 49  
Cancer, 23  
Code génétique, 38  
Coopérativité enzymatique, 96  
Cryomicroscopie électronique, 91, 96, 107  
Cycle de la sélénocystéine, 47  
Décamères protéiques, 106, 241  
Discrimination entre S et Se, 20  
Discrimination par SelA, 97, 98  
Incorporation cotraductionnelle, 35  
Locus et opérons, 53, 229  
Mécanisme réactionnel, 94, 103  
Matériels et Méthodes, 245  
Phosphosérine (Sep), 56, 100  
Phylogénie, 55, 69, 75, 86, 99, 229, 232, 234  
Pyridoxal-5'-phosphate (PLP), 92  
Pyrrolysine (Pyl), 51  
Recodage, 43  
Ribosome, 35, 44, 46  
Sélénium, 18, 61  
Sélénocystéine (Sec), 33, 47, 60  
Sélénocystéine synthase (SelA), 41, 83, 90  
Sélénophosphate (SePO<sub>3</sub><sup>3-</sup>), 31, 41, 65, 87  
Sélénophosphate synthétase (SelD / SPS),  
41  
Sélénophosphate synthétases (SelD / SPS),  
31, 65, 89  
Sélénoprotéines, 26, 63  
SECIS, 45  
SelA-like archaebactériennes, 104  
Sep-ARN<sub>t</sub><sup>Cys</sup> Cys-ARN<sub>t</sub><sup>Cys</sup> synthase (SepCysS),  
55, 99  
Sep-ARN<sub>t</sub><sup>Sec</sup> Sec-ARN<sub>t</sub><sup>Sec</sup> synthase (SepSecS),  
41, 100  
Stœchiométrie chez SelA, 94  
Toxicité du sélénium, 19  
Traduction protéique, 35, 44, 46  
Usages industriels du sélénium, 19  
Végétaux supérieurs et sélénium, 21





---

# Abstract

## Structural and enzymatic study of the complex between tRNA<sup>Sec</sup> and selenocysteine synthase SelA.

Selenium is both a toxic pollutant and an essential trace element, which is contained in selenocysteine, known as the 21<sup>st</sup> amino acid. In selenoproteins, this amino acid is often a primordial residue situated in the enzymatic active site. The incorporation of selenocysteine is performed by the ribosome during translation and involves recoding of a Stop codon (UGA) of the selenoprotein's messenger RNA. The mechanism requires a signal sequence on this mRNA called SECIS (SElenoCysteine Insertion Sequence). Thus, selenocysteine is an exception of the genetic code : unlike canonical amino acids, it is synthesised on its tRNA, and subsequently incorporated by specific factors. This original biosynthesis on the tRNA<sup>Sec</sup> is found, in slightly different forms, in all three domains of life.

In Eubacteria, SelA is the enzyme responsible for the selenocysteine biosynthesis on its tRNA<sup>Sec</sup>. We report here the structural and enzymatic studies of the complex between SelA and tRNA<sup>Sec</sup>, for the organisms *E. coli* and *M. thermoacetica* taken as examples. This report presents first the crystallography of both complexes followed by a biochemical study of the macro-complex formation between decameric SelA and up to 5 tRNA<sup>Sec</sup>. We took advantages of techniques such as microcalorimetry, chemical footprinting and activity measurement. Finally, a discussion on these new data towards the understanding of SelA function and origin is described.

**Keywords:** *genetic code, protein-RNA complex, tRNA<sup>Sec</sup>, SelA, selenium, selenocysteine.*

## Résumé

Le sélénium est à la fois un polluant toxique et un oligo-élément essentiel, qui entre dans la composition de la sélénocystéine : le 21<sup>ème</sup> acide aminé. Dans les sélénoprotéines, cet acide aminé est souvent un résidu primordial situé au site actif de l'enzyme. L'incorporation de la sélénocystéine est effectuée lors de la traduction, par le ribosome, grâce à un "recodage" d'un codon Stop (UGA) de l'ARN messager de la sélénoprotéine. Le mécanisme requiert la présence d'une séquence signal dans cet ARNm appelée SECIS (SEleno- Cysteine Insertion Sequence). La sélénocystéine est donc une exception du code génétique : contrairement aux acides aminés canoniques, elle est synthétisée sur son ARN<sub>t</sub>, puis incorporée grâce à des facteurs spécifiques. Cette biosynthèse originale sur l'ARN<sub>t</sub><sup>Sec</sup> se retrouve, sous des formes légèrement différentes, dans les trois domaines du vivant.

Chez les eubactéries, l'acteur de l'étape clé de cette biosynthèse est SelA, enzyme responsable de l'incorporation du sélénium conduisant à la sélénocystéine au niveau de l'ARN<sub>t</sub><sup>Sec</sup>. En prenant comme exemples les organismes *E. coli* et *M. thermoacetica*, cette thèse rapporte l'étude structurale et enzymatique du complexe formé par SelA et l'ARN<sub>t</sub><sup>Sec</sup>. Il est abordé dans un premier temps la cristallisation du complexe pour les deux organismes d'étude, puis l'étude biochimique de la formation du macro-complexe entre l'enzyme décamerique SelA et jusqu'à 5 ARN<sub>t</sub><sup>Sec</sup>. Nous avons pour cela tiré parti des techniques de micro-calorimétrie, d'interférences chimiques et de mesures d'activité. Une discussion sur l'apport de ces nouvelles données sur l'interprétation du fonctionnement du complexe complète cette étude.

**Mots-clés:** *ARN<sub>t</sub><sup>Sec</sup>, code génétique, complexe protéine-ARN, SelA, sélénium, sélénocystéine.*