Universidad Autónoma de Madrid

Facultad de Ciencias
Departamento de Biología Molecular

Doctoral Thesis

Genes and proteins involved in RNA modification: evolutionary genomic context and characterization of YibK and GidB Methyltransferases

Alfonso Benítez Páez

June 2010

Director
Dr. María Eugenia Armengod

Dedicatoria

Esta tesis no ha sido fruto de un esfuerzo individual sino de todas y cada una de las personas que estuvieron, están y seguirán conmigo el resto de mi vida. Porque fueron muchos los que me apoyaron y ayudaron a conseguir este logro personal y profesional lleno de dificultades e inconvenientes. A todas esas personas que estuvieron a mí alrededor brindandome su apoyo en los mejores y peores momentos les dedico este trabajo. También esta dedicado a aquellos que confiaron en mi y que creyeron que sacaría esto adelante. A mis padres por haberme dado la oportunidad de estudiar, ellos fueron los responsables del empujon inicial. A mis hermanos por hacerme sentir enormemente valorado y demostrarme lo orgullosos que estan de mí. Para el resto de mi familia por brindarme cariño y confianza. A Anahi, Leo y Virna, que han sido mi familia desde mi llegada a España, gracias por sus consejos, apoyo, cariño, ayuda, fortaleza y sobretodo por vuestra sabiduria. A Ana y Julia por su inmensa ayuda, incondicional apoyo y eterno cariño que me fueron necesarios en los últimos momentos. Finalmente, quiero dedicar este logro a Juan Esteban y Sonia, tal vez ellos son los principales autores de este trabajo porque sin el apoyo de ellos no hubiera podido culminar mi doctorado. Aunque durante todo este tiempo existieron

numerosos momentos en los que nuestras vidas se tornaron complicadas, siempre existio el deseo de tener un futuro lleno de oportunidades y enseñanzas para nuestro hijo. Simplemente dedico este trabajo a ellos porque desde hoy en adelante se merecen recoger el fruto de todos los sacrificios hechos en los últimos años.

El hombre inteligente no es el
que tiene muchas ideas,
sino el que sabe sacar provecho
de las pocas que tiene.

CONTENT

Acknowledgements

I wish to thank the entire Molecular Genetics group headed by María Eugenia Armengod for its support during my PhD career, thanks guys (Magda, Ismï, Ana, Rodri, Silvia, Rafa, Mª Jose, Natalia, Elvira, and Carmen) for the huge number of hours of dedication, discussions, teaching and friendship. I also wish to thank the Cellular Biology Group (Inma, Carmen, Nuria, Felix, Jose, Ghita, Juan, Raja, Asun, and Erwin) for its support and for those funny crazy chats and yelling days inside our lab. Special thanks to Leonardo Arbiza for his support and encouraging me each time I was down; our nights of chats and beers were so useful to reflect on myself. Thanks to Anahi for her unconditional support. Thanks also go to Virna for her care from faraway lands on the other side of Mediterranean Sea. I am so grateful to Toni Gabaldon for his academic support and teaching. Special thanks to Ana and Julia for their invaluable help and care. Thanks to Isabel Cárdenas, Dávide Báu and Sabina Mekong for their friendship and support. I want to thank my whole family for its love, support and for taking pride in me. Thanks to my parents, Hernan and Maria, my brothers, Luis and William, my sister, Laura, and my dear aunt, Esther. I'll be eternally grateful for all the love and support from Sonia and my dearly beloved son, Juan Esteban, who were the inspiration to hold this dream. THANKS!

Overview

The post-transcriptional modification of ribonucleotides is a common feature in the non-coding RNAs of the three major phylogenetic domains of life: Archea, Bacteria and Eukarya. The density of modified nucleotides per molecule is higher in tRNA (transfer RNA) than in rRNA (ribosomal RNA). Based on *Escherichia coli* data, rRNA has a density of 1 modification per hundred ribonucleotides, whereas tRNA has approximately 1 modification per 10 nucleotides. This distribution of modifications suggests they play a more critical role in the functions accomplished by tRNAs. However, different studies have demonstrated that the absence/presence of some modifications in rRNA is able to cause profound effects on cell physiology such as aminoglycoside resistance and translation failure, and may even impair ribosome assembly. Consequently, they are localised in critical regions for decoding mRNA. In tRNAs, post-transcriptional modifications are well-known for being critical for not only their decoding functions, but their L-shape structure maintenance, aminoacylation and stability.

Nowadays, more than one hundred different modifications have been identified and found in the RNA of the above-mentioned life domains. However, the number of characterised RNA-modifying genes is not fully depicted. In light of this, there is a moderate set of orphan modifications

whose responsible genes are still expected to be identified, especially for modifications in tRNA.

To date, 31 different modifications can be found in the tRNAs of *Escherichia coli*, and approximately 30 genes have been characterised to be involved either directly and indirectly in tRNA modification. Although we can assume that most tRNA-modifying genes are known, we must take into account that a large number of modifications is not composed of simple steps of modification, methylation for instance. Thus, several modifications comprise multi-step reactions involving more than one enzyme (i.e., the IscS/TusA complex to incorporate a "thio" group into position 2' of Uridine 34 in some tRNAs; or the MnmE/MnmG complex to produce the wobble cmnm^{5}U34 modification in some tRNAs). Thus, it is easy to expect a full set of tRNA-modifying genes to be higher than the modifications observed.

Unlike in tRNA, modifications in both bacterial 16S and 23S rRNAs are less frequent along these molecules and are simpler modifications than those harboured in tRNAs. Thanks to the different genomics and the domain, motif, structural and architecture information of the genes and proteins stored in biological databases, several studies based on comparative genomics have favoured the characterisation of rRNA-modifying enzymes, which have been rapidly discovered in the last few years. Consequently, the coverage of the genes known to be responsible for rRNA modifications in *Escherichia coli* is 86%.

Nowadays, the characterisation of RNA-modifying genes is a promising field to study because protein translation, and consequently bacterial growth, could be severely affected when RNA-modifying enzymes are deleted. Hence, a potential antimicrobial therapy based on enzyme targeting may prove worthwhile since some pathogenic effects in bacteria can be considerably weakened by the loss of RNA modifications. Accordingly, modifications in RNA seem to work as a mechanism to control the expression of a specific set of genes in the cell. Therefore, it is plausible to think that RNA modifications are able to decode the multiple and hidden molecular signatures encoded in the cell genome which remain undetectable at the sequence level. For this very reason, a direct co-evolution of genomes and translation machinery, including RNA modifications and RNA-modifying genes, could be expected.

Different approaches and strategies are being used to determine the effect of several modifications in decoding, and how they affect translation fidelity. Despite several translation failures evidenced when some tRNA and rRNA modifications are lacking, the evolutionary meaning of the post-transcriptional modifications emerging as molecular epitopes to tune the codon·anticodon pairing in order to properly decode a genome has not been fully explained.

The aim of this thesis is to study the RNA modifications with different perspectives with a view to gaining a better understanding of their molecular role. We outlined three different objectives in this thesis: i) to describe some expression features of an operon encoding RNA-modifying genes; ii) to study the functional and biochemical aspects of *E. coli* RsmG methyltransferase acting on rRNA 16S; and iii) to search novel tRNA-modifying genes using computational approaches based on comparative genomics. The most important results of our research have enabled us to know both the new and different regulatory signals acting at the transcriptional and translational levels to control the expression of the MnmG and RsmG proteins. Simultaneously, we disclose a considerable set of amino acids that are critical for the *in vivo* function of RsmG. The residues where the co-factor binding, catalysis, and RNA binding tasks lie were pinpointed by sequence conservation and structural localisation in the protein. Additionally, effects on translation were detected in the *ΔrsmG* mutant, showing an error-prone phenotype in the read-through and frameshifting assays.

Finally, a comparative genomics approach was applied to find the new tRNA-modifying genes directly involved in the synthesis of orphan tRNA modifications to date. By applying this strategy, we retrieved relevant information about the evolution among the different components of translation machinery. At the same time, several of the genes which are probably involved in tRNA modification were detected according to the evolutionary interactions predicted. Different experimental strategies were applied to investigate the possible targets for some candidate genes. Based on the structural and functional information of our candidate proteins, we studied the modification status of the $tRNA^{Leu}_{CmAA}$ and $tRNA^{Leu}_{cmnm5UmAA}$ in the *ΔyfiF* and *ΔyibK* mutants. These tRNAs carry the wobble modifications Cm and cmnm^{5}Um, respectively. Using the RNA mass spectrometry

approach, we found that *yibK* is responsible for the Cm and Um modifications in these tRNAs. As a result, we show a reliable methodology to predict the molecular function of genes according to their genomic context. In order to contribute to the knowledge of the new genes involved in tRNA modification, we also present this methodology openly for it to be applied in the search for the missing components of other relevant pathways in the cell.

Besides the results presented in this thesis, we wish to make known general concepts of RNA modification biology in the best comprehensible way. We expect all the readings found inside the various chapters of this book to be user-friendly for everyone familiar with biology. The concepts offered herein not only represent the individual effort made to investigate the intricate field of translation emphasised in the meaning of RNA modifications, but are the result of a researching career of a multidisciplinary group which practices basic science and focuses on studying one of the most relevant molecular processes to support the complex process of life.

Resumen

Las modificaciones post-transcripcionales de nucleótidos son una característica predominante en los RNAs no codificantes (aquellos RNAs que no se traducen en polipétidos) presentes en todos los organismos pertenecientes a los tres principales grupos taxonómicos: Arquea, Bacterias y Eucariotas. Al hablar de las modificaciones en RNA no codificante debemos esencialemente enfocarnos en aquellas presentes en moléculas como el tRNA (RNA de transferencia) y el rRNA (RNA de los ribosomas), las cuales durante muchos años han sido estudiadas y enmarcadas en un papel relevante para el correcto funcionamiento de la maquinaria de traducción de la célula. Durante el estudio de las modificaciones del RNA se ha podido establecer que ellas también son fundamentales para la estabilidad estructural de los mismos. En el caso de rRNAs, la ausencia de ciertas modificaciones esencialmente se ha asociado con la resistencia a antibióticos. Por otra parte, la carencia de modificaciones en tRNAs se ha asociado con la pérdida de estabilidad estructural de los mismos, fallos en el proceso de aminoacilación y pérdida de la fidelidad de lectura en el ribosoma, el cual se vuelve propenso a cometer errores en la descodificación del mRNA.

A día de hoy se conocen más de un centenar de nucleósidos modificados, los cuales se distribuyen general o particularmente en los diferentes grupos

taxonómicos. No obstante, algunas de esas modificaciones permanecen sin una asociación directa con la enzima responsable de su síntesis.

En *Escherichia coli* han sido identificados más de 30 genes responsables de modificaciones encontradas en sus tRNAs. Aunque podría asumirse que el grupo de genes responsables de las 31 modificaciones encontradas en sus tRNAs esta casi determinado, se debe tener en cuenta que muchas de esas modificaciones son llevadas a cabo en varias reacciones y por más de una enzima participante (ej. el complejo IscS/TusA que incorpora el grupo "thio" en la posición 2' de la Uridina 34 de varios tRNAs; o el complejo MnmE/MnmG que produce el nucleósido cmnm^{5}U34 también en varios tRNAs). De esa forma se puede esperar que el número de genes responsables de modificaciones en tRNAs sea mayor que las modificaciones en si mismas.

Ya sea en rRNA o tRNA, el estudio de la influencia de las modificaciones de RNA es un campo prometedor debido a que la traducción de proteínas y consecuentemente el crecimiento celular se ven seriamente afectados cuando determinados genes responsables de modificaciones en el RNA son delecionados en diferentes modelos bacterianos. Con lo cual, posibles efectos antimicrobianos podrían ser esperados al inhibir la actividad de ciertas enzimas modificadoras. La indispensable presencia de modificaciones, esencialmente en tRNAs, ha sido extensamente documentada demostrando su relevante papel en patogenicidad, virulencia, respuesta a cambios de pH y respuesta a cambios de temperatura en diversos organismos. A partir de esos datos, es fácilmente concebible que las modificaciones en el RNA podrían actuar a modo de mecanismo de regulación, controlando la expresión de varios grupos de genes necesarios bajo ciertas condiciones fisiológicas en la célula.

La era genómica ha traido consigo información concerniente a la presencia y distribución de genes modificadores de RNA en diferentes organismos, así como el conocimiento de motivos moleculares y estructurales de proteínas modificadoras. Todo este tipo de información ha servido para que en los últimos años se hayan identificado un elevado número de enzimas encargadas de modificar RNA.

El objetivo principal de esta tesis es el estudio de las modificaciones que ocurren en el RNA desde diferentes perspectivas con el fin de aportar

información que ayude al entendimiento de su papel dentro de la descodificación del genoma. De esta manera, se han trazado tres objetivos principales: i) describir el patrón de expresión del operón *gid*, el cuál codifica los genes *mnmG* y *rsmG* implicados en modificación de tRNA y rRNA, respectivamente; ii) estudiar bioquímica y funcionalmente la metiltransferasa RsmG de *Escherichia coli* que modifica el rRNA 16S; y iii) la búsqueda de nuevos genes implicados en la modificación de tRNA mediante el uso de estrategias biocomputacionales basadas en la genómica comparativa.

Entre los principales resultados obtenidos podemos destacar el hallazgo de nuevos elementos reguladores de la expresión génica dentro del operón *gid* y que afectan directamente la expresión de *rsmG.* Por otra parte, se pudo analizar y comprobar la relevancia funcional de un amplio grupo de residuos necesarios para la correcta función de la enzima RsmG. Mediante diferentes aproximaciónes experimentales se pudieron establecer residuos implicados en la unión del cofactor AdoMet, residuos implicados en la catálisis así como aquellos posiblemente involucrados en la unión al RNA. Todos ellos pudieron ser evidenciados gracias a estrategias computacionales basadas en la conservación de secuencia y análisis estructurales de la misma enzima.

Finalmente, y a través de estrategias usadas en genómica comparativa, hemos querido realizar un estudio a gran escala para el hallazgo de nuevos genes implicados en las modificaciones de tRNA en *E. coli*. Gracias a esta estrategia pudimos establecer un interesante patrón de co-evolución entre muchos elementos participantes en la traducción de proteínas, entre ellos, las enzimas de modificación de tRNA y rRNA. De tal manera, se pudo recuperar un grupo de genes de función desconocida cuya participación en procesos de modificación es muy probable dado su contexto genómico y los fenotipos asociados con fallos en la traducción que pudieron ser apreciados en su abordaje experimental. Después de un análisis funcional de dominios se pudieron establecer posibles dianas para algunos de los 11 genes candidatos. Consecuentemente, en los mutantes nulos de los genes *yibK* y *yfiF,* que codifican enzimas de actividad metiltransferasa de tipo SPOUT, se estudio la presencia de las modificaciones cmnm5Um y Cm, presentes en la posición 34 de $tRNA^{Leu}_{cmnm5UmAA}$ y $tRNA^{Leu}_{CmAA}$, respectivamente. Mediante el estudio de tRNA a través de espectrometría de masas se pudo establecer que el gen *yibK* codifica una proteína responsable de la metilación de tipo 2'-*O* en la ribosa

del ribonucleótido ubicado en la posición de tambaleo (posición 34) de dichos tRNAs.

Además de los resultados expuestos en los siguientes apartados de esta tesis, queremos sobre todo divulgar los conceptos generales de la biología de la modifcación del RNA de la mejor manera posible para su correcto entendimiento. Esperamos que los conceptos expresados sean fáciles de comprender para todos aquellos familiarizados con la biología. De igual forma, como autor de este trabajo, quiero declarar que los conceptos presentados aquí no solo representan un esfuerzo individual para investigar este complejo proceso dentro de la traducción de proteínas, sino que también son el resultado de los esfuerzos comunes de un grupo de investigación multidisciplinar dedicado al estudio y comprensión de uno de los procesos moleculares más relevantes de la célula e indispensable para la vida.

INTRODUCTION

The post-transcriptional modification of nucleotides is a common process occurring in non-coding RNAs (ncRNA) such as ribosomal RNA (rRNA) and transfer tRNA (tRNA). The critical role of these modifications has been evidenced through several experimental studies that support their involvement in the proper reading of the genetic information encoded in mRNAs. Given that rRNAs and tRNAs are primary players of ribosomal architecture, their incomplete post-transcriptional processing is frequently associated with ribosome misreading failures.

The phylogenetic distribution of RNA modifications also sheds light on their relevance for translation. Some tRNA modifications are highly conserved in all three major kingdoms of life: Archea, Bacteria and Eukarya. In this way, the wide distribution of modifications supports the notion of an ancient crucial role in translation. Notwithstanding, this conservation of modified nucleotides could be achieved by either divergent evolution from a common ancestor or convergent evolution (Björk & Hagervall, 2005b).

Given that the main scope of this thesis is to study aspects of the RNA modification of the rRNAs and tRNAs of *Escherichia coli*, the next sections refer to the state of the art, essentially in this model organism.

Modified Nucleosides

A wide variety of modified nucleosides is found in both rRNA and tRNA molecules. However, most of the hypermodified forms or baroque alterations of common A, G, U, and C nucleosides reside in this last type of ncRNAs. All the modifications are confined to the addition of simple or complex chemical groups to the purine or pyrimidine rings and to the 2'-hydroxyl group of ribose moiety of nucleotides. In this last case, incorporation of a simple methyl (CH_3) group, and not another type of modification, is predominant.

Having observed the diversity and complexity of modifications, a consensus nomenclature was established to refer to the individual modifications occurring in the different nucleotides.

- First of all, different chemical groups can be added to the classical A, U, G, and C nucleosides. In this case, common lowercase abbreviations to the left of the modified nucleoside, followed by an exponent, refer to the position where substitution was achieved. Then, abbreviations like c, i, k, m, n, o, r, s and t correspond to the frequent carbonyl, isopentenyl, lysyl, methyl, amino, oxy, ribosyl, thio and threonyl groups, respectively (Björk & Hagervall, 2005b). Thus, an adenosine, which is methylated in position 2 of its purine moiety, is represented as m^2A.
- In addition, other abbreviations such as hU, Ψ, I, Q, and GluQ refer to dihydrouridine, pseudouridine, inosine, queuosine, and glutamylqueuosine, respectively. These other types of nucleosides can undergo modifications, as well as the classical A, G, U, and C. In this way, methylations in the Ψ nucleosides are frequently found in rRNA (i.e., $m^3\Psi$).
- The modifications occurring in ribose moiety (primarily methylations) are denoted with an "m" to the right of the modified nucleoside (i.e., Cm, Gm, or Um).
- Finally, the location of the modification is delimited with a number corresponding to the nucleotide's position in the RNA sequence. Then m^2G1516 and S^2C32 are the full nomenclatures for the common modified nucleosides located in position 1516 of rRNA 16S and in position 32 of the tRNAs of *Escherichia coli,* respectively (Björk & Hagervall, 2005b; Ofengand & Campo, 2005b).

To date, over one hundred modifications have been found in the ncRNAs belonging to archea, bacteria and eukaryotes (the RNA modification Database at http://library.med.utah.edu/cgi-bin/rnafind.cgi). However, most of them are still orphan, which means that the enzymes responsible for them are unknown. Most of the modifications appearing in the rRNA and tRNA molecules have been both chemically and spatially mapped. When modifications are compared between both types of RNAs, a difference in terms of both number and complexity is easily noted.

A higher density of modified nucleosides in tRNAs is clearly evident. Approximately 10% of tRNA nucleosides are modified, as opposed to only 1% of rRNA nucleosides. Furthermore, a difference in the nature of modifications is also visible. In rRNA, modifications are primarily restricted to: i) isomerisation of U to Ψ; ii) addition of methyl groups to purine and pyrimidine rings; and iii) addition of methyl groups to the 2'-hydroxyl ribose moiety of the nucleoside (Decatur & Fournier, 2002; Ofengand & Campo, 2005b). tRNAs show a wide spectrum of modifications that are characterised by the incorporation of bulky groups, especially into both the anticodon and the adjacent anticodon regions (i.e., cmnm^{5}Um34 in the tRNA^{Leu} of *Escherichia coli*) (Figure 1).

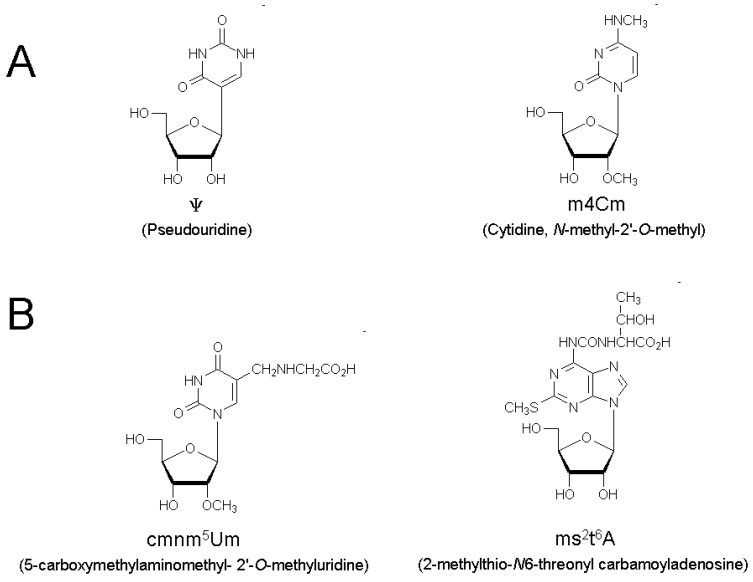

Figure 1. Structure of the different nucleosides found in rRNAs and tRNAs. A – Common modifications found in rRNA molecules. B – Hypermodified nucleosides found in tRNAs.

Modifications in Ribosomal RNAs

As stated above, modifications in rRNAs are less frequent than in tRNAs. Approximately 1 per hundred nucleotides is modified, and this density is observed in both the 16S and 23S molecules of *Escherichia coli*. However, the lower distribution of modified nucleotides is compensated by their biased distribution where most modifications spatially converge in the respective three-dimensional structure of the ribosome subunits (discussed below).

The main feature of rRNA modifications is the poor diversity of the modifications they show. The isomerisation of uridines to pseudouridines and the simple addition of methyl groups to nucleotides are predominantly found in rRNAs. Notwithstanding, they have diverse effects which are essentially associated with changes in the nucleoside electrostatic charge. In this way, all the methylations, except m^7G, enhance local hydrophobicity, while the conversion of U into Ψ promotes an additional hydrophilic H-bond formation (Ofengand & Campo, 2005b). In the *E. coli* rRNA 16S, 11 different modifications can be found and the set of modifications is extended to 25 in rRNA 23S. Their sequence and structural localisation are depicted in Figures 2 and 3, respectively.

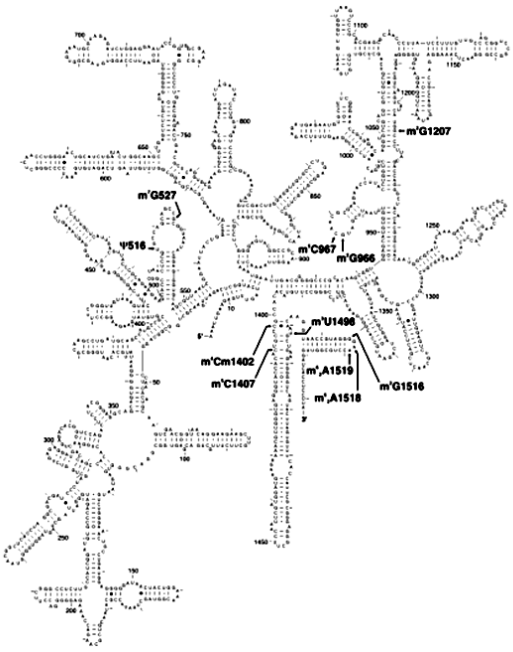

Figure 2. Distribution of the modified nucleoside in the rRNA 16S of *Escherichia coli* [from (Ofengand & Campo, 2005b).

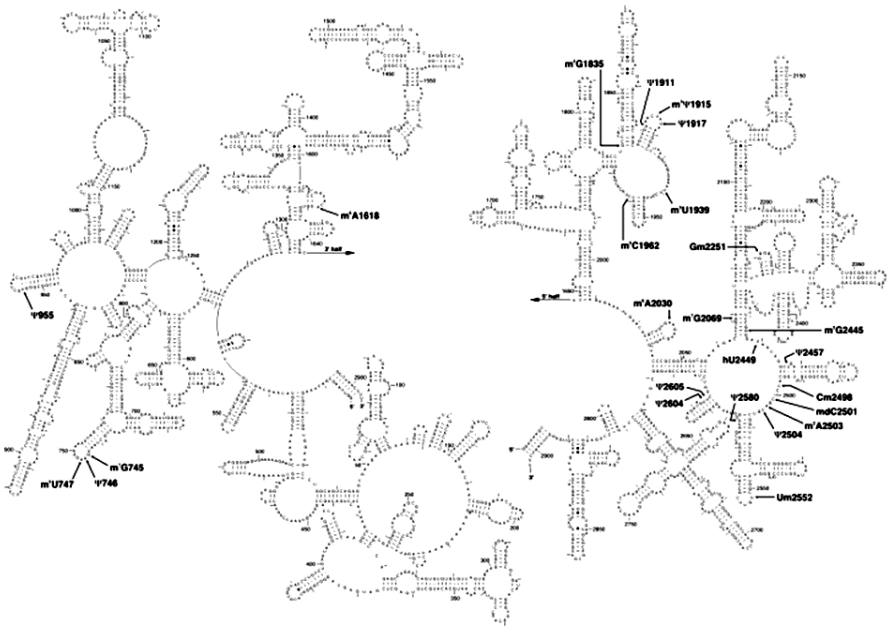

Figure 3. Distribution of the modified nucleosides in the 23S rRNA of *Escherichia coli* [from (Ofengand & Campo, 2005b)].

Nowadays, most of the enzymes responsible for rRNA modifications have been characterised. Moreover, the search for rRNA-modifying genes has become a promising scenario thanks to several computational approaches based on comparative genomics that have been used to identify rRNA methyltransferases by the sequence, structure and the genomic information extracted from several bacterial genomes currently available in public databases (Andersen & Douthwaite, 2006; Basturea et al., 2006; Lesnyak et al., 2006; Sergiev et al., 2006; Purta et al., 2008a; Purta et al., 2008b; Sergiev et al., 2008; Purta et al., 2009).

To date, just 5 modifications of the entire set of rRNA modifications found in *Escherichia coli* are still orphan (Table 1). Those modifications are: m^2G1516 in rRNA 16S, and m6A2030, m7G2069, hU2449 and mdC2501 in rRNA 23S.

Pseudouridines

Pseudouridine (Ψ) was not only the first to be discovered, but the most abundant modified nucleoside in rRNA. It is a product of the uridine isomerisation that confers new physicochemical properties to Ψ to benefit

the rRNA structure and function. The chemical differences between the U and Ψ nucleosides are shown in Figure 4.

Table 1. List of the modified nucleosides present in the 16S and 23S rRNAs of *E. coli* [from (Purta et al., 2009).

		Modifying enzyme		
Nucleotide	Modification	Name	Synonym	Reference
16S rRNA				
516	Ψ	RsuA	YejD	Wrzesinski *et al.* (1995a)
527	m^7G	RsmG	GidB	Okamoto *et al.* (2007)
966	m^2G	RsmD	YhhF	Lesnyak *et al.* (2007)
967	m^5C	RsmB	Fmu, YhdB	Tscherne *et al.* (1999a); Gu *et al.* (1999)
1207	m^2G	RsmC	YjjT	Tscherne *et al.* (1999b)
1402	m^4Cm	RsmH[a,b]; RsmI[a,b]		
1407	m^5C	RsmF	YebU	Andersen and Douthwaite (2006)
1498	m^3U	RsmE	YggJ	Basturea *et al.* (2006)
1516	m^2G	RsmJ[b]		
1518	$m^6{}_2A$	RsmA	KsgA	Helser *et al.* (1972); Poldermans *et al.* (1979)
1519	$m^6{}_2A$	RsmA	KsgA	Helser *et al.* (1972); Poldermans *et al.* (1979)
23S rRNA				
745	m^1G	RlmA[I]	RrmA,YebH	Gustafsson and Persson (1998)
746	Ψ	RluA	YabO	Wrzesinski *et al.* (1995b)
747	m^5U	RlmC	YbjF, RumB	Madsen *et al.* (2003)
955	Ψ	RluC	YceC	Conrad *et al.* (1998); Huang *et al.* (1998)
1618	m^6A	RlmF	YbiN	Sergiev *et al.* (2008)
1835	m^2G	RlmG	YgjO	Sergiev *et al.* (2006)
1911	Ψ	RluD	Yfil	Huang *et al.* (1998); Raychaudhuri *et al.* (1998)
1915	$m^3Ψ$	RluD	Yfil	Huang *et al.* (1998); Raychaudhuri *et al.* (1998)
		RlmH	YbeA	Purta *et al.* (2008a); Ero *et al.* (2008)
1917	Ψ	RluD	Yfil	Huang *et al.* (1998); Raychaudhuri *et al.* (1998)
1939	m^5U	RlmD	YgcA, RumA	Agarwalla *et al.* (2002); Madsen *et al.* (2003)
1962	m^5C	RlmI	YccW	Purta *et al.* (2008b)
2030	m^6A	RlmJ[b]		
2069	m^7G	RlmK[b]		
2251	Gm	RlmB	YjfH	Lovgren and Wikstrom (2001)
2445	m^2G	RlmL	YcbY	Lesnyak *et al.* (2006)
2449	D	RldA[b]		
2457	Ψ	RluE	YmfC	Del Campo *et al.* (2001)
2498	Cm	RlmM	YgdE	This study
2501	*C[c]	RltA[b]		
2503	m^2A	RlmN	YfgB	Toh *et al.* (2008)
2504	Ψ	RluC	YceC	Conrad *et al.* (1998); Huang *et al.* (1998)
2552	Um	RlmE	FtsJ, RrmJ	Caldas *et al.* (2000); Bugl *et al.* (2000)
2580	Ψ	RluC	YceC	Conrad *et al.* (1998); Huang *et al.* (1998)
2604	Ψ	RluF	YjbC	Del Campo *et al.* (2001)
2605	Ψ	RluB	YciL	Del Campo *et al.* (2001)

Ψ is the only modified nucleoside to exhibit a C-C rather than the common N-C bond that links base and sugar, which confers greater functional flexibility than U. Besides, the additional free N1-H atom promotes a new H-bond formation (see Figure 4). As a result of the above features, Ψ reveals novel pairing abilities in rRNA (Charette & Gray, 2000). At the same time, Ψ enhances the local base stacking in both the single and duplex RNA regions thanks to the induction of the 3'-endo conformation of ribose which restricts the pyrimidine ring to an axial *anti* conformation. Consequently, the stacking effect extended to the neighbour nucleotides in RNA is propagated through helical regions, conferring a global stability to the RNA structure (reviewed in (Charette & Gray, 2000)).

Pseudouridine synthases

In *Escherichia coli,* there are seven different Ψ synthases which are able to originate all the 11 Ψ present in both 16S and 23S rRNAs (Ofengand & Campo, 2005b). This correlation between enzymes and modifications results from the capacity of RluC and RluD to synthesise 6 different Ψ, 3 of each (see Table 2).

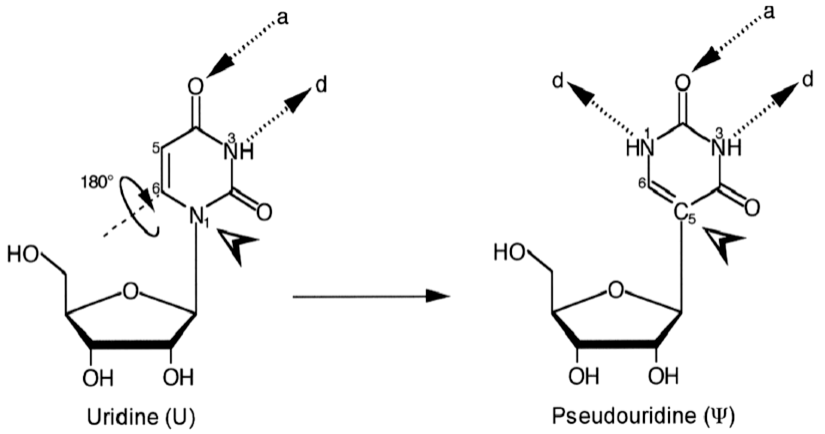

Figure 4. The chemical differences between the U and Ψ nucleosides [from (Charette & Gray, 2000)]. An additional H-bond donor is shown in the dotted arrow "d" in the Ψ N1 atom.

However, there are no different synthases acting on the same U nucleotides to produce an equivalent Ψ. As Table 2 shows, all the rRNA Ψ synthases are grouped into two major families of proteins: RsuA and RluA. Together the TruA and TruB families of proteins acting on tRNAs, rRNA Ψ synthases, are considered an ancient superfamily of proteins with a low overall amino acid sequence identity. Notwithstanding, they share very short conserved motifs which allow one to hypothesise that they have all possibly emerged from multiple duplications of an ancestral Ψ synthase (Charette & Gray, 2000; Ofengand & Campo, 2005b).

Despite all the families of Ψ synthases exhibiting a low sequence identity, this class of proteins structurally converges (Del Campo et al., 2004; Kaya et al., 2004). Furthermore, some members of this family of proteins tend to contain ancestral RNA binding domains, such as PUA and S4, fused to the main Ψ synthase domain (Anantharaman et al., 2002a).

Table 2. The pseudouridines present in the rRNAs of *E. coli* [from (Ofengand & Campo, 2005b).

Ψ site	Synthase	Previous gene name	*E. coli* Swiss-Prot	*S. enterica* Swiss-Prot	Family
16S rRNA					
516	RsuA	*yejD*	P33918	Q8XGP8	RsuA
23S rRNA					
746	RluA	*yabO*	P39219	Q8ZRV9	RluA
955	RluC	*yceC*	P23851	Q8ZQ16	RluA
1911	RluD	*yfiI, sfhB*	P33643	Q8XGG2	RluA
1915	RluD	*yfiI, sfhB*	P33643	Q8XGG2	RluA
1917	RluD	*yfiI, sfhB*	P33643	Q8XGG2	RluA
2457	RluE	*ymfC*	P75966	Q8XPZ1	RsuA
2504	RluC	*yceC*	P23851	Q8ZQ16	RluA
2580	RluC	*yceC*	P23851	Q8ZQ16	RluA
2604	RluF	*yjbC*	P32684	Q8ZKL1	RsuA
2605	RluB	*yciL*	P37765	Q8ZP51	RsuA

On the other hand, the catalytic mechanism for the synthesis of Ψ has been extensively studied. Structural data combined with residue conservation have led to a likely hypothesis in which a highly conserved Asp in all the families of the Ψ synthases (motifs GRLD and HRLD of the RsuA and RluA families, respectively) would mediate the isomerisation of U by a nucleophilic attack on the C_6 atom of the uracil ring. Nonetheless, another mechanism is still viable and is supported by the attack of C_1' of ribose moiety (reviewed in (Ofengand & Campo, 2005b)).

METHYLATIONS

Escherichia coli rRNAs have 2-fold more methylated nucleosides than Ψ. The entire set of methylations occurring in *E. coli* rRNAs is shown in Table 1. Although previously argued that modifications in rRNA are less frequent than tRNA, they are not randomly distributed in the structure (see Figures 2 and 3). Therefore, this biased distribution implies that modifications strongly influence both the structure and function of ribosomes. The methylations in rRNA nucleotides are restricted and occur in either the ring of the base or the 2'-hydroxyl group of ribose moiety, and both types of methylations confer hydrophobicity to the nucleoside where the methyl group is incorporated (Decatur & Fournier, 2002; Ofengand & Campo, 2005b).

A tendency to localise where specific translation events take place is easily observed when the three-dimensional distribution of methylations (as well as Ψ) is seen (Figure 5). They are oriented towards the interaction faces of the ribosome subunits. Despite methylations conferring a hydrophobic property

to RNA, which possibly permits an interaction with other ribosome components such as proteins, the absence of modifications in those regions dominated by ribosomal proteins indicates that the RNA-protein interaction is not mediated by modifications (Decatur & Fournier, 2002). In *E. coli* 16S rRNA, most methylations are distributed close to the mRNA channel at the P and A sites. Therefore, they must act on the maintenance of the proper reading frame during translation (Decatur & Fournier, 2002). Direct evidence of rRNA methylations being involved in critical translation steps has been recently shown. In this way, the 16S rRNA modifications, m^2G1207 and m^3U1498, carried out by the RsmC and RsmE methyltransferases (MTases), respectively, were found to be directly implicated in the translation initiation by recognising the anticodon stem of the initiator $tRNA^{fMet}$ (Das et al., 2008). Moreover, the influence of other methylations in translational fidelity is also known (van Buul et al., 1984; O'Connor et al., 1997).

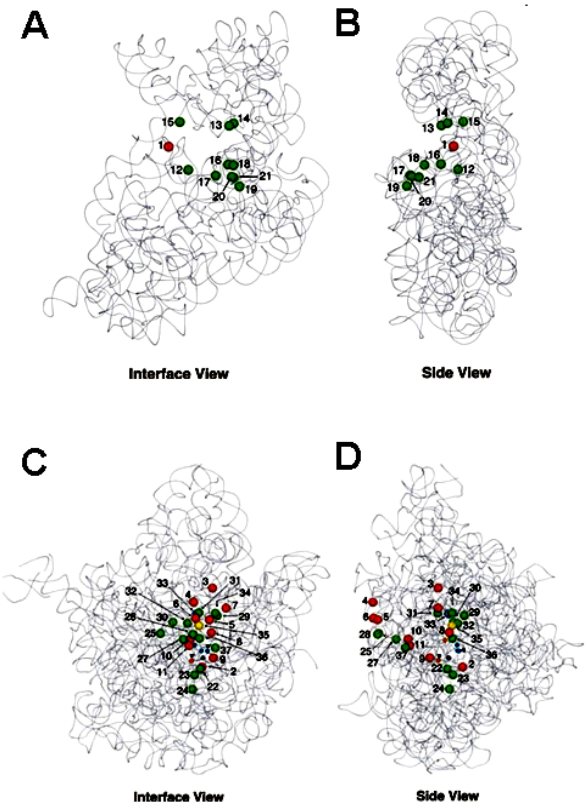

Figure 5. Three-dimensional localisation of the modifications in the *E. coli* rRNAs. Panels A and B represent the different views of 16S rRNA, while panels C and D show views of 23S rRNA. Red circles show the localisation of Pseudouridines. Green circles depict the localisation of methylations [from (Ofengand & Campo, 2005b)].

Despite the fact that the absence of rRNA methylations apparently does not have profound effects on bacterial growth or survival, another relevant aspect with which rRNA methylations have been associated is antibiotic resistance. Johansen and co-workers demonstrated that the loss of the respective Cm1409 and Cm1920 modifications in *Mycobacterium tuberculosis* rRNA 16S and 23S confers resistance to Capreomycin and Viomycin. Simultaneously, this fact was corroborated when a susceptibility to these aminoglycosides was ascertained through the expression of recombinant *M. tuberculosis tlyA* (responsible for those modifications) in *E. coli* (Johansen et al., 2006). Other antibiotic resistances, such as kasugamycin and streptomycin, have also been associated with the lack of rRNA modifications as accomplished by genes *rsmA* and *rsmG*, respectively (Zimmermann et al., 1973; Okamoto et al., 2007). In short, it is assumed that a lack of rRNA methylations impairs aminoglycoside binding at these sites.

The ribosomal RNA methyltransferases

In global terms, the enzymes that methylate RNA comprise two major classes of MTases based on their structure core: i) the Rossmann-Fold methylases (RFM), which include almost all the N and C methylases and modify the RNA bases, and ii) the superfamily of RNA MTases SPOUT, which consists in the 2'-O-methylases relating to TrmD and SpoU, both of which essentially act in tRNAs (Anantharaman et al., 2002a, b).

Notwithstanding, a later classification of MTases redistributed them as five structurally distinct classes of MTases, which were denoted as I (RFM), II, III, IV (SPOUT) and V (reviewed in (Schubert et al., 2003). Interestingly, a global sequence conservation among all the MTases classes is poor, but they structurally manifest an analogous architecture as a result of the functional convergence to use S-adenosyl-L-methionine (AdoMet) as a cofactor of the methyltransfer reaction (Martin & McMillan, 2002; Schubert et al., 2003). Class I Mtases comprises most rRNA-modifying enzymes (and DNA MTases), which show a fair degree of sequence similarity in agreement with the reaction they perform (Ofengand & Campo, 2005b).

As noteworthy features, the consensus G-x-G-x-G (G-X-G, at least) AdoMet-binding motif is highly conserved in most MTases, even in the most structurally dissimilar Mtases classes (classes III, IV and V). At the same time, the low degree of sequence similarity observed in the predominantly Class I

MTases acting on RNA does not clarify the evolutionary history of these proteins. Therefore, multiple independent lineages may explain the predominance of this class of MTases for RNA modification. However, functional evolutionary convergence is not fully discarded since the global sequence comparison supports no evident relationships (Schubert et al., 2003). Most rRNA MTases are well conserved only in bacteria and cannot be traced in other kingdoms such as eukarya. Nevertheless, a few genes (i.e., *rsmA*) have a wide phylogenetic distribution that confers these conserved MTases a relevant role in decoding both the function and biogenesis of the ribosome (Anantharaman et al., 2002a; O'Farrell et al., 2008; Xu et al., 2008).

Some examples of these conserved rRNA MTases, the RsmA and RsmG (known as GidB) enzymes, are mentioned. RsmA is an evolutionary, well conserved protein responsible for both the m^6_2A1518 and m^6_21519 modifications in 16S rRNA (these modifications are also found in the small ribosomal RNA of the mitochondrion in eukaryotes) (Zimmermann et al., 1973). Of the full set of MTases that act on rRNAs, RsmA was the first to be characterised and, unlike others, the RsmA/Dim1 family has been the most studied by far. In addition to its role in rRNA modification, RsmA is assumed to be involved in additional tasks given its association with the cold-sensitive suppression phenotype and the acid-shock response noted in the *rsmA* mutants (Lu & Inouye, 1998; Inoue et al., 2007). The other well conserved MTase RsmG is responsible for m^7G527 methylation in the same rRNA 16S. Unlike RsmA, RsmG has not been further studied. Consequently, our aim is to study the different functional aspects of this enzyme and its resulting modification to determine its role in both cell physiology and the decoding process of this conserved modification.

Modifications in Transfer RNA

As mentioned earlier, tRNAs undergo greater modifications than rRNAs. Consequently, it is easy to find that at least 10% of the nucleotides in tRNAs have been modified. Nonetheless, the percentage of modified nucleotides in tRNAs can be largely extended in other organisms. For instance in eukarya, it is possible to find up to 25% of modified nucleotides and, globally, this higher frequency of modifications in tRNA could correspond to a plethora of interactions in which tRNAs are involved (Björk & Hagervall, 2005b).

Not only is the complexity of tRNA modification higher, but also its frequency. Although the major modifications distinguished in rRNA can also be observed in tRNA (i.e., methylations and Ψ), several voluminous groups are frequently linked to nucleotides in this last type of RNA (see Figure 1). Such ornate modifications sometimes comprise the incorporation of intact amino acids, which also occurs in t^6A, k^2C and GluQ modifications where the threonyl, lysis, and glutamyl groups are respectively added to the modified nucleotides.

The presence of modifications in tRNA is of ancient origin, and even though most of them come to light in specific phylogenetic groups, a few tRNA modifications are present as a well conserved phenomenon among the three major kingdoms (see Figure 6).

tRNA modifications are distributed along the sequence and they play different roles in translation. Mainly, as mentioned above in previous sections, the post-transcriptional modifications in tRNA also help stabilise a particular tertiary structure and are, therefore, required for the proper functioning of this class of RNA (Davis, 1998). tRNA modifications also help to modify the cognate codon recognition in order to affect the aminoacylation and to stabilise the codon-anticodon wobble base pairing to prevent ribosomal frameshifting (reviewed in (Björk & Hagervall, 2005b)).

The study of tRNA modifications has essentially centred on investigating their role in the decoding function. Therefore, modifications taking place in the wobble position (commonly known as position 34 of tRNAs) and in the adjacent anticodon regions (positions 32 and 37) are extensively examined to explain their molecular role in mRNA reading and in their evolutionary meaning.

In this context, tRNA modifications are considered to influence translation and regulate gene expression by improving the decoding capabilities of tRNA, which directly affects codon sensitivity, codon choice and maintains the reading frame (Bjork & Rasmuson, 1998). Furthermore, given that tRNA modification results in a complex process involving the multiple reactions, enzymes and tRNAs, various degrees of modification can be seen in the different physiological states of the cell (Emilsson & Kurland, 1990; Bjork & Rasmuson, 1998). As a result, the role of tRNA-modifying enzymes in maintaining the modified status of tRNAs would be mediated

directly by their expression. In this context, the regulatory signals controlling the operons where tRNA-modifying genes are encoded play a pivotal role in the modification of tRNAs, where the major regulation of RNA modifications would be based on an unbalance between enzymes and substrates (Winkler, 1998). For this reason, it would be very useful to study the regulatory signals of the operons containing tRNA-modifying genes to investigate the levels of expression and response to the different states or stress conditions in the cell.

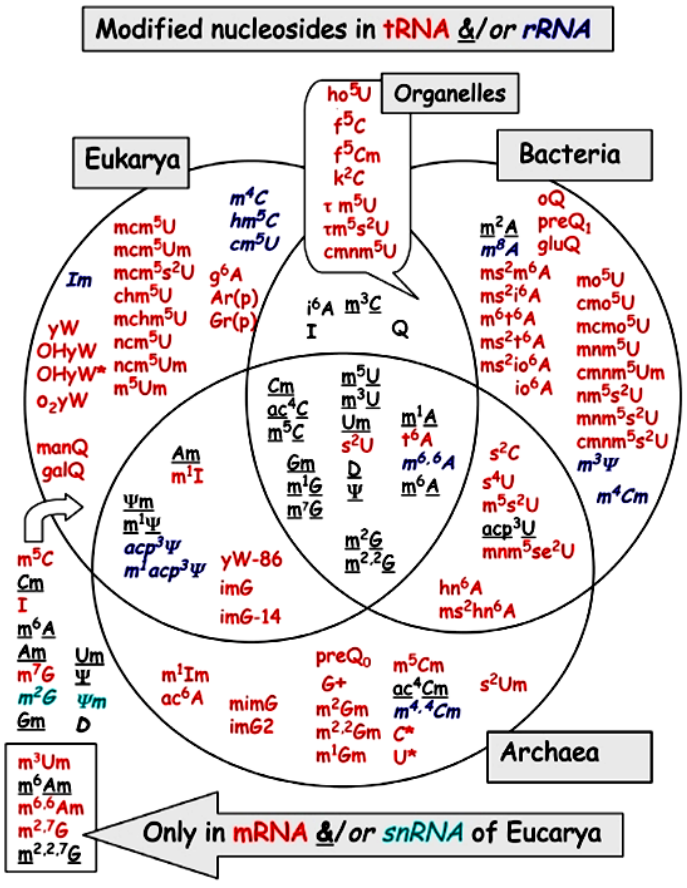

Figure 6. The phylogenetic distribution of the nucleosides modified in RNA [from (Grosjean, 2009)]

tRNA MODIFICATIONS IN THE WOBBLE POSITION

Position 34 of tRNA, usually termed as "wobble", shows the greater miscellany of modifications seen for any other position in tRNAs (Figure 7).

The wobble position pair with the third codon base (referred to as N_3 from this point onwards): i) is the only one permiting non-Watson-Crick interactions in the codon·anticodon pairing, ii) affects the affinity for codons by altering the anticodon loop structure, and iii) expands or restricts the decoding capabilities of tRNAs (Curran, 1998).

Originally, Francis Crick's wobble hypothesis proposed that U34 could pair with A_3 or G_3, but not with U_3 or C_3 (Crick, 1966). Nowadays, it is known that a large amount of the modifications of tRNAs taking place in position 34 extend the ability of this nucleotide to pair with others (reviewed in (Agris et al., 2007)). This property has been experimentally demonstrated in those tRNAs harbouring xo^5U modifications in the wobble uridine (Nasvall et al., 2004, 2007). In constrast, other modifications restrict wobbling to decode the specific purine, or pyrimidine-ending codons, of the mixed boxes.

Among the modifications that restrict wobbling, we see that uridine-2-thio-5-carboxymethylaminomethyl ($cmnm^5S^2U34$) is found in the tRNAs that read the CAA and CAG codons for Gln, as well as the UUA and UUG codons for Leu, and the uridine-2-thio-5-methylaminomethyl (mnm^5S^2U34) is found in the tRNAs decoding the Lys, Glu and Gly codons (see Figure 8). These modifications are characteristic of the tRNAs decoding all the mixed codon boxes and they permit the reading of A/G-ending codons, thus preventing missense errors (Björk & Hagervall, 2005a). The role of these modified tRNAs to decode mRNA has been efficiently tested in translational assays, demonstrating that wobble modifications are critical for translation fidelity (Elseviers et al., 1984; Hagervall & Bjork, 1984b; Bregeon et al., 2001).

In addition to the mixed codon boxes which contain the two-fold degenerated codons, there are four-fold degenerated codons in the genetic code as well (known as family codon boxes). Such codons are read by the tRNAs for Ala, Val, Pro, Thr, Leu and Ser (see Figure 8). In *Escherichia coli,* these family codon boxes are decoded by tRNAs with xo^5U derivates in position 34, such as the uridine-5-oxyacetic acid (cmo^5U34) modification or its methylester form ($mcmo^5U34$). Although a reviewed version of Crick's hypothesis argues that a modified U34 would be able to pair with A_3, G_3 and U_3, but not with C_3 (Okada et al., 1979), it has been recently demonstrated that cmo^5U34 is able to pair with all the different nucleotides at the third position of the family codon boxes (Nasvall et al., 2004, 2007). Accordingly, these modifications extend the capability of tRNAs to read all four variants of

the family codon boxes, whereas the wobble modifications of the tRNAs decoding two-fold degenerated codons restrict the reading of the A/G- or U/C-ending codons (Björk & Hagervall, 2005b). This evidence highlights the evolutionary meaning of the tRNA modifications emerging as molecular epitopes to strengthen, promote, and tune codon·anticodon pairing to properly decode the genome.

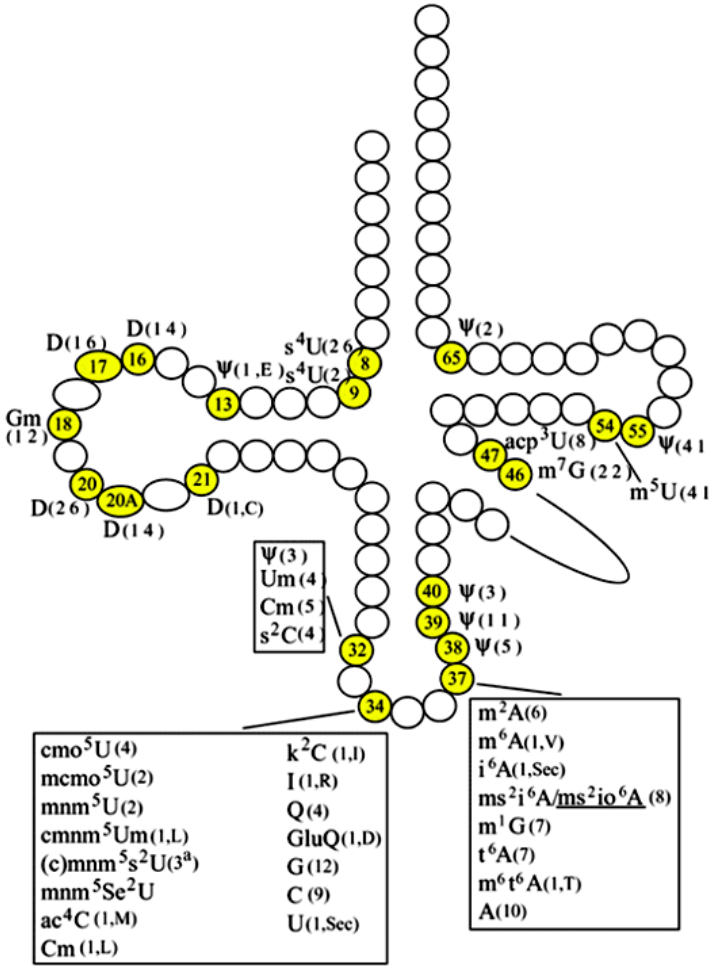

Figure 7. Localisation of the modified nucleosides in tRNA. The yellow positions are those in which modifications are present. The numbers in parentheses indicate the number of tRNA species where the respective modification is present (in some cases, a one-letter amino acid code is present) [from (Björk & Hagervall, 2005b)].

Other wobble modifications comprise the inosine (I) nucleoside resulting from a deaminated adenosine. I34 has a wide phylogenetic distribution in all three major kingdoms of life (Rozenski et al., 1999). Bacteria have a single

I34-containing tRNA that reads the arginine codons CGA, CGC, and CGU (see Figure 8). In eukaryotes, I34-containing tRNAs decode various family codon boxes and the isoleucine AUU, AUC, and AUA codons (Curran, 1998). Accordingly, I34 appears as a modification which extends the decoding properties of tRNAs as well as the cmo^5U wobble modification. The null presence of A34 rather than I34 in tRNAs exhibits a response in computational studies which conclude that I34 improves wobble base-pairing and, unlike A34, does not weaken the initial ribosomal A-site binding when at the P-site (reviewed in (Agris et al., 2007)).

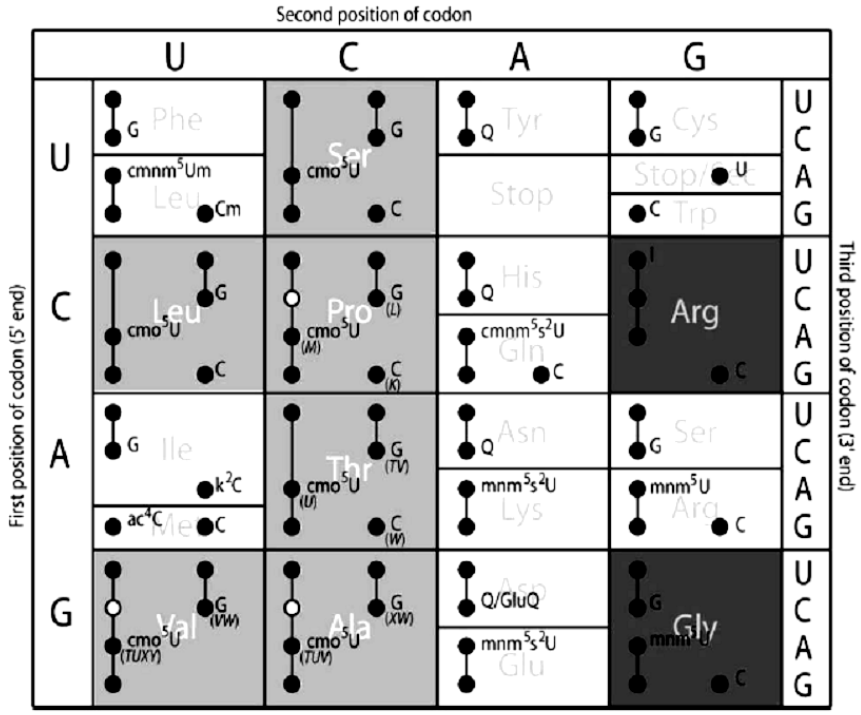

Figure 8. The distribution of wobble modifications in the tRNAs of *E. coli* and *S. enterica*. The decoding properties per tRNAs modified are also shown [from (Nasvall et al., 2007)].

Queuosine (Q) is a frequently found modified nucleotide in the tRNAs decoding the mixed codon boxes for Tyr, His, Asn, and Asp (see Figure 8). Given its distribution in tRNAs, the role of Q is to ensure the correct reading of the U/C-ending codons in these boxes. This ability to avoid interaction with the A/G-ending codons could be conceived through a *syn* conformation, which inhibits the Q:R (R = purine) pair (reviewed in (Curran, 1998)). By using the *his* operon however, known to be regulated by attenuation

(Johnston et al., 1980), those mutants deficient in *tgt* (the gene responsible for Q synthesis) show no difference in the *his* expression compared to that seen in wild-type cells, indicating that Q-lacking $tRNAs^{His}$ efficiently decode the His codons in the leader region of the *his* attenuator. Consequently, Q34 does not seem to influence the decoding efficiency of tRNAs, but has been seen to be involved in the efficient binding of $tRNA^{Tyr}$ to ribosomes by decreasing tRNA binding by 2-fold when Q34 is lacking (reviewed in (Björk & Hagervall, 2005b).

Nucleotide ac^4C (N4-acetylcytidine) is present only in the elongator $tRNA^{Met}$. Early studies using the $tRNA^{Met}CAU$ lacking the ac^4C modification indicate that aminoacylation was not affected when C34 remained unmodified. Subsequently, it has been demonstrated that $tRNA^{Met}$ without the ac^4C modification allows tRNA to misread Ile AUA and to increase the efficiency to recognise the complementary AUG codon. Globally, these results support that the function of ac^4C nucleosides is to prevent a misreading of the AUA codon read by the minor species $tRNA^{Ile}_{k2CAU}$, which is achieved by decreasing Met AUG codon affinity. Therefore, this last tRNA is favoured in competition for AUA (Stern & Schulman, 1977, 1978; Björk & Hagervall, 2005b).

As mentioned above, the k^2C (lysidine) modification is present in the minor tRNA species reading the AUA Ile codon (Muramatsu et al., 1988). The synthesis of k^2C34 is carried out by the *tilS* product, which is essential for cell viability (Soma et al., 2003). k^2C34 prevents any misacylation of tRNA with methionine instead of isoleucine, and synergistically confers fidelity to AUA, which reads the ac^4C34 modification, as explained in the previous paragraph (Soma et al., 2003; Björk & Hagervall, 2005b).

Modifications in position 32

Position 32 of tRNAs is less modified than positions 34 or 37. However, it retains a greater variety of modifications than the positions outside the anticodon region. It is the first nucleotide of the seven positions of the anticodon loop (see Figure 7). A pyrimidine is always found in position 32, followed by a universal U nucleotide in position 33. The role of the modification taking place at position 32 is essential to confer stability to the anticodon loop structure through the H-bond formation with the nucleotide in position 38 (Baumann et al., 1985; Björk & Hagervall, 2005b). Four different modifications occur in position 32 in the tRNAs of *Escherichia coli*

(see Figure 7), and their involvement in translation fidelity has been documented for at least the S^2C modification. Thus, S^2C seems to play a role in the prevention of ribosomal frameshifting (Baumann et al., 1985).

Modifications in position 37

Like the wobble position, position 37 of tRNA is largely modified and shows conserved rules that tolerate the presence of certain nucleotides in this position. Here, only purines are allowed and an A is predominantly observed that is frequently modified with large groups. The presence of G is less frequent, even though it is preferably methylated (see Figure 7). Globally, nucleotide 37 is conditioned to retain a very hydrophobic status that confers the anticodon region with both structural and functional properties for the purpose of proper mRNA decoding.

Although this is a modified nucleotide that does not directly interact with any position of the codon, it plays a relevant role in translation efficiency. Consequently, different degrees of modification correlate with the N36·N_1 interaction type (the third position of the tRNA anticodon and the first position of the codon in mRNA). Therefore, a weak A·U interaction requires the presence of the nucleotides modified with bulky groups, such as ms^2i^6A37 or t^6A37, which grant additional hydrophobicity to A37 in order to stabilise the weak N36 interaction with N_1 in the codon. Consequently, a stronger N36·N_1 interaction based on the G·C pair requires more simple modifications like m^2A37 or m^6A37 (Björk & Hagervall, 2005b). Moreover, structural roles have also been associated with modifications in position 37. In this way, the modifications incorporated into base 37 avoid N33·N37 pairing and, at the same time, promote the anticodon loop from stacking. In addition, modifications in position 37 prevent frameshifting, improve cell growth and influence the ternary complex formation rate (reviewed in (Björk & Hagervall, 2005b)). Basically, modifications in position 37 have profound effects on cell physiology as they control translation efficiency.

Modifications outside the anticodon region

Cell growth shows no critical dependence on the lack of modifications occurring outside the anticodon regions. Although it has been previously discussed that modifications, such as Ψ, act as a structure stabiliser, the fact they are lacking has minor effects on both growth and ribosomal reading (Björk & Hagervall, 2005b).

tRNA-modifying enzymes

Besides the pseudouridine synthases and the class I and IV MTases responsible for simple methylations in tRNA, the set of enzymes participating in tRNA modification comprises several families of proteins with diverse contents of both the structural and functional domains.

The well characterised SPOUT (class IV) superfamily of MTases is traditionally associated with post-transcriptional RNA modification given the role of its members in either tRNA and rRNA processing (Anantharaman et al., 2002b; Purta et al., 2006; Tkaczuk et al., 2007; Purta et al., 2009). This family of MTases has specific structural features that differ from the classical MTase fold. One of those features is their ability to dimerise. In this way, the cofactor binding close to the "knot" (the reason why some of them are frequently called TREFOIL knot proteins) of one monomer is stabilised by dimerisation (reviewed in (Tkaczuk et al., 2007)). Furthermore, many SPOUT members contain additional RNA binding domains fused to the C- or N-terminal ends. Some of these domains are S4 (found in S4 ribosomal protein), PUA (from Pseudouridine synthase and Archaeosine transglycosylase) (Aravind & Koonin, 1999), TRAM (Anantharaman et al., 2001), THUMP (Aravind & Koonin, 2001) or OB-fold (Murzin, 1993).

Albeit an extensive duplication in evolution is thought to produce the several families of RNA methylases known, it cannot be ruled out that multiple lineages of RNA MTases could emerge as convergent evolution given the poor global sequence similarity they share (Anantharaman et al., 2002a).

Base thiolation is carried out by a well studied group of enzymes involved in many tasks such as protein-protein interaction, assembly and transference of the Fe-S cluster and, finally, RNA interaction. Some of them exhibit fusions to RNA binding domains (i.e., TRAM or THUMP), and resemble the SPOUT MTases. Additionally, some enzymes appear to have a metal cluster-containing domain that catalyses sulphur insertion (reviewed in (Anantharaman et al., 2002a)). The sulphur used to incorporate nucleotides is derived from cysteine, and the IscS protein is the cornerstone in this process. In addition, *iscS* co-transcribes with several other genes directly involved in Fe-S cluster formation (reviewed in (Leipuviene et al., 2004; Björk & Hagervall, 2005b). The assembly and mobilisation of Fe-S cluster requires an intricate process in which specific sets of proteins are independently involved

to provide sulphur groups to thiolate different tRNA positions (see Figure 9). Normally, a cystein desulphurase (primarily IscS), a Fe-S cluster assembly/scaffold protein and, finally, an RNA-binding protein that incorporates the sulphur group into tRNA, are all required to accomplish thiolation in different tRNA positions. (Nilsson et al., 2002; Björk & Hagervall, 2005b; Lundgren & Bjork, 2006).

NTPases, that is, proteins which hydrolyse both ATP and GTP, are the most ancient components of translation machinery. It is well-known that they act as translation factors to permit initiation and elongation. The evolution of major GTPase lineages has produced new classes of proteins which are involved in specific functions within translation as a whole. For instance, the Era family contains a known RNA binding domain, KH, a pseudoKH domain, and similar motifs corresponding to the G-domain, which is also found in the well conserved TrmE protein (Anantharaman et al., 2002a). In this way, a relevant role is expected for this uncharacterised family of proteins given its surprising architecture and phylogenetic distribution. A clear involvement of the GTPases in tRNA modification has been demonstrated for MnmE. The *mnmE* mutant was early associated with the lack of the common mnm^{5}U group present in the wobble position of several tRNAs in bacteria (Elseviers et al., 1984). Subsequently, the achievement of the xm^{5}U modifications has been shown to depend on GTPase activity (Yim et al., 2003; Martinez-Vicente et al., 2005). Nowadays, it is well-known that xm^{5}U modifications synergistically depend on the MnmE and MnmG (previously called GidA) proteins (Bregeon et al., 2001; Yim et al., 2006; Moukadiri et al., 2009).

Together with MnmE, the MnmG family of proteins is responsible for cmnm^{5}U synthesis and its derivatives which may be found in the wobble position of the tRNAs reading the A/G-ending codons of mixed boxes (see Figure 8). MnmG is a well conserved protein that is present in all the phylogenetic kingdoms, and it also belongs to the FAD-binding superfamily of proteins with an unclear NAD(H)-binding domain and a vast surface for tRNA anchoraging (Meyer et al., 2008; Osawa et al., 2009a; Osawa et al., 2009b). Besides its orthologues, paralogue members of this family have also been detected. Accordingly, shorter versions of the MnmG protein (~650 aa) can be found in organisms such as *Myxococcus xanthus* (~450 aa) (White et al., 2001) or *Thermus thermophillus* (~230 aa) (Iwasaki et al., 2004). Nonetheless, the shorter version of MnmG, called TrmFO, is involved in

frequent m^5U54 methylation using the N^5,N^{10}-methylenetetrahydrofolate as cofactor (Urbonavicius et al., 2005). MnmA (also known as TrmU), MnmG and MnmE, together with the MnmC (with the last one in a minor way because of its restricted distribution to bacteria) proteins, make up an interesting puzzle of functional domains that are needed to produce a significant modification in the wobble positions in tRNAs (see Figure 10). Given the different proteins and their respective functional domains which are involved in such modifications, and finally their high degree of evolutionary conservation, it is coherent to believe that the modification they synthesise is highly critical to decode the genome during evolution in a consensual and unclear way.

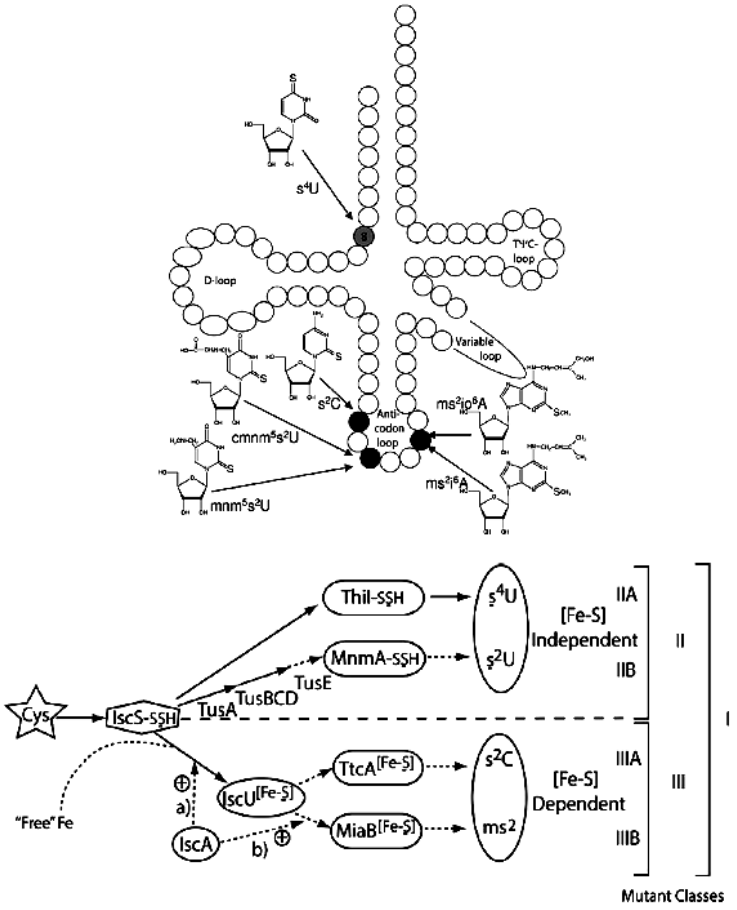

Figure 9. Thiolated nucleotides in the tRNA of *S. enterica*. The sequence and structural localisation of the thiolated nucleotides in tRNA are shown at the top of the figure. Different pathways of thiolation are depicted at the bottom of the figure [from (Lundgren & Bjork, 2006)].

Another class of proteins implicated in tRNA modification, and one that has been less studied, is dihydrouridine synthases. This type of proteins also belongs to the FAD-binding superfamily and can be found in all three primary kingdoms given the broad distribution of this modified nucleoside in RNAs. As with the other proteins acting as tRNA-modifying enzymes, dihydrouridine synthases are frequently fused to RNA binding domains, for example, LRP1 Zn-finger, dsRBD and CCCH (Anantharaman et al., 2002a). Evolutionary studies of these enzymes suggest that they emerged early in bacterial evolution and were subsequently transferred to eukaryotes, probably by endosymbiosis. Notwithstanding, the diversification of dihydrouridine synthases occurred independently in bacteria and eukaryotes (reviewed in (Anantharaman et al., 2002a)).

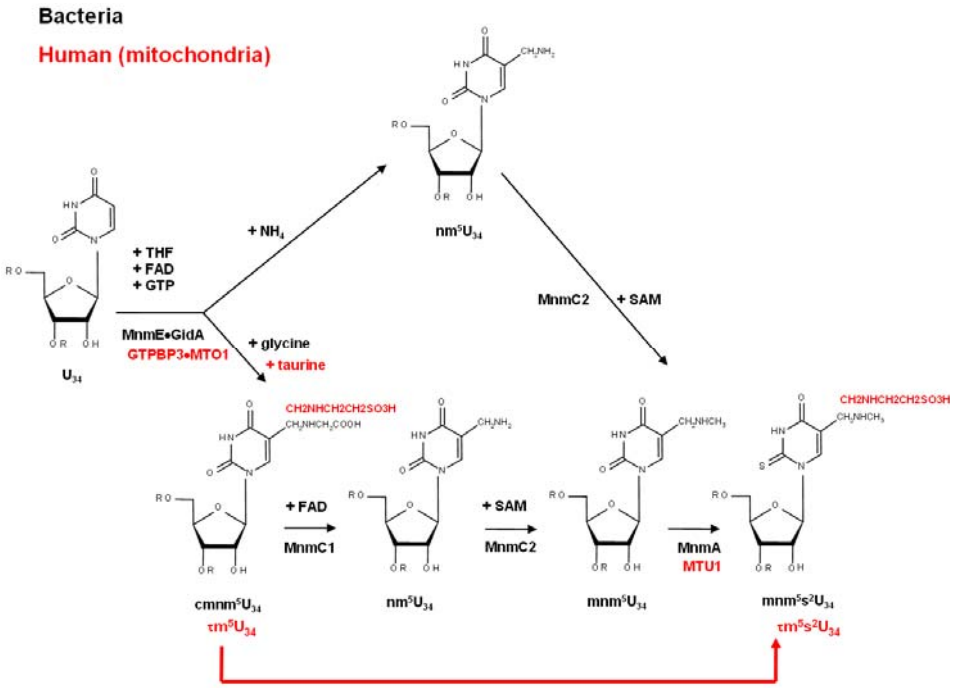

Figure 10. The biosynthetic pathway controlled by the conserved MnmE and MnmG proteins [adapted from (Moukadiri et al., 2009)].

Globally, tRNA-modifying proteins have emerged as multi-domain entities that are capable of recognising both sequence and structural epitopes on tRNA by fusing with well-known RNA binding domains. Despite their closeness by the reactions they achieve, they independently evolve to grant specificity to the modification process which consequently confers stability and reading properties to the tRNA Modified.

Despite the great effort made to characterise those enzymes responsible for tRNA modifications, some modified nucleosides still await the characterisation of those enzymes which promote their synthesis. Among this set of orphan modifications we find the acp^3U present in position 47 of several tRNAs. Other unexplored modifications are of specific interest as they lie in the anticodon region of tRNAs. These modifications are the 2'-O-methylations present in the wobble positions of $tRNA^{Leu}_{cmnm5UmAA}$ and $tRNA^{Leu}_{CmAA}$, the C2-methylations present in A37 of some tRNAs, and the N6-methylation present in the t^6A37 of $tRNA^{Thr}_{GGU}$. Given this scenario, the objective of this study is to also search for new tRNA-modifying proteins using different computational and experimental approaches.

Objectives

1. Describing the expression features of the genes encoding a couple of proteins involved in RNA modification that conform a highly conserved operon among bacteria. This analysis will lead to valuable information about the expression pattern of the MnmG and RsmG proteins in the cell. The analyses will consider both transcriptional and translational regulation aspects.

2. Biochemically and functionally characterising the RsmG protein of *Escherichia coli* implicated in the synthesis of the conserved modified nucleotide m^7G527 at rRNA 16S. Experimental approaches will be conducted to corroborate a prior sequence and structure of *in silico* analyses of this family of proteins.

3. Searching for novel tRNA-modifying proteins through comparative genomic approaches. The analysis of the genomic context of known tRNA-modifying enzymes in multiple bacteria and eukaryote genomes will provide information about the co-evolution patterns among the multiple components of translation machinery.

Results

Chapter 1
Expression Analysis of the *Escherichia coli gid* operon

Summary

The enzyme-dependent chemical modifications of rRNA and tRNA have been demonstrated as being critical for decoding mRNA in the cell. One of the mechanisms controlling this translational issue is regulated by the expression of RNA-modifying genes. The gid operon comprises the *gidA* (mnmG) and *gidB* (rsmG) genes, which are directly involved in the modifications occurring in tRNA and rRNA, respectively. Although previous studies have presented data on the gidA expression in relation to the transcriptional activity in *oriC*, there is no further information about its expression pattern during cell growth nor information about its possible coupled expression with *gidB*. Here, we aim to study certain transcriptional and translational features of the *gid* operon in order to correlate the information obtained with the functional significance of the RNA modifications they perform. Interestingly, we have found unknown transcriptional elements that directly influence the *gidB* expression in the operon. Moreover, we discuss the role that these elements play in properly

controlling the *gidB* expression to guarantee the necessary protein levels to modify rRNA. Finally, we found a correlation between the *gid* gene expression and that already known for its RNA substrates.

INTRODUCTION

The *gid* operon is assumed to form one transcriptional unit, including the *gidA* and *gidB* genes, in most bacterial genomes. Interestingly, the *gid* operon contiguously maps at the replication origin in the *Escherichia coli* chromosome. This localisation is an evolutionarily conserved feature in several bacterial groups. Due to its localisation, and the experimental analyses in which the expression pattern of *gidA* correlates with the replication activity in *oriC* (Theisen et al., 1993; Bogan & Helmstetter, 1997), both genes were associated with cell division early on. According to previous analyses, the *gidA* expression is up-regulated early in cell growth. However, its transcription is briefly repressed after the initiation of replication. This expression pattern could be involved in preventing the premature triggering of chromosome replication (Ogawa & Okazaki, 1994). The *gid* genes have been cited because a growth-inhibited phenotype was observed in the *E. coli gidA* mutants in glucose media (*glucose-inhibited division* genes) (von Meyenburg et al., 1982). In von Meyenburg's study, a phenotype with a modified *gidB* expression was also detected after *gidA* disruption. Therefore, it is assumed that *gidA* and *gidB* form a transcriptional unit. However, no studies have further characterised the transcriptional activity in both genes.

Transcriptional studies into *gidAp* (*gidA* promoter) have shown a specific inhibition of the expression by ppGpp (Ogawa & Okazaki, 1991). This inhibition is similar to that known for the tRNA and rRNA genes after triggering the stringent control during amino acid starvation.

Unlike *gidA*, the *gidB* expression has not yet been studied, and the *gidB* expression is expected to have the same pattern as *gidA* given its closeness. A hypothetical coupled expression of these genes could be also explained by their molecular role given that they are both involved in RNA modification.

In contrast to a priori functionality of the *gid* genes inferred from physical mapping, various studies published in the last few years have revealed that *gidA* and *gidB* functions are not directly associated with the cell division

process. g*idA* (now called *mnmG*) has been described as a gene that encodes a 70-kDa enzyme involved in tRNA modification. GidA activity is associated with MnmE, which is a GTPase involved in the same modification pathway of tRNA biosynthesis (Bregeon et al., 2001; Scrima et al., 2005; Yim et al., 2006; Moukadiri et al., 2009). *gidA* mutants show decoding failures as evidenced in the phenotypes of translational misreading (Bregeon et al., 2001). The GidA protein is well conserved among bacteria and eukaryotes and belongs to a large family of FAD-binding proteins.

While GidA acts as a tRNA modification protein, GidB was characterised as a rRNA modification enzyme a few years ago (Okamoto et al., 2007). In *Escherichia coli*, *gidB* (currently known as *rsmG*) codes a SAM-dependent methyltransferase that is responsible for the m^7G527 modification in rRNA 16S. Although the *gidB* mutants seem to have no delayed growth rate as the *gidA* ones do, the lack of m^7G527 modification in rRNA 16S is associated with the low-level streptomycin resistance in many bacteria where this mutation has been studied (Nishimura et al., 2007a; Nishimura et al., 2007b; Okamoto et al., 2007). Another interesting and unexplained phenotype of the *gidB* mutants is the emerging frequency of high-level streptomycin resistance mutants which is at least 200 times greater than that observed for the wild-type ones (Okamoto et al., 2007).

gidB is present only in bacterial genomes. Notwithstanding, *gidB*-like versions can be detected in some eukaryote genomes, mainly in the Viridiplantae clade. Those plant homologues are easily retrieved when a simple Blast comparison of the GidB protein from *Escherichia coli* against the non-redundant protein database of plants is done. Consequently, identity values of 38% and 31% for the GidB-like proteins coded in *Oryza sativa* and *Arabidopsis thaliana,* respectively, are found.

The clustered arrangement of *gidA* and *gidB* suggest that their expression would be coupled in an operon manner. However, no further experimental evidence has been published to date. Here we show a set of results disclosing the transcriptional and translational features of the *gid* operon. g*id* genes have been confirmed to be expressed in an operon fashion. Nevertheless, we describe a specific transcriptional regulation for *gidB* in which almost 75% of the transcripts generated from *gidAp* do not reach *gidB*. Additionally, we describe a new promoter able to provide the further expression of *gidB* alone.

Furthermore, the expression characterisation at the protein level has demonstrated that GidB has a lower half-life than that determined for GidA.

METHODS

STRAINS AND PLASMIDS

The different strains and plasmids used in this study are listed in Table 1. The *gidA::kan* and *gidB::kan* mutants were kindly donated by the National BioResource Project (NIG, Japan). The clones identified as 4, 8, and 10 carrying the *gidA:Tn10* mutation were obtained from D. Bregeon and coworkers (Bregeon et al., 2001). All the mutations were recovered in a Dev16 background derived through P1 procedures (Miller, 1990). The correct insertion of mutations was checked by genomic PCR using primers that are specific for insertion cassettes and flanking genomic regions. The MC4100 strains and their rpoS::Tn10 mutant were kindly donated by Dr. Miguel Vicente at the Centro Nacional de Biotecnología – CNB in Madrid, Spain.

Table 1. List of the strains and plasmids used in this study.

Id	Description	Source
	Strains	
IC4639	Dev16 *lacZ105*(amb), derivated from Elseviers 1984 (IC4639)	Yim, 2003
IC5678	BW25113 *gidB::Kan* (*lacI*q, *rrnB*T14, Δ*lacZ*WJ16, *hsdR*514, Δ*araBAD*AH33, Δ*rhaBAD*LD78)	NBRP, Japón
IC5831	BW25113 *gidA::Kan* (*lacI*q, *rrnB*T14, Δ*lacZ*WJ16, *hsdR*514, Δ*araBAD*AH33, Δ*rhaBAD*LD78)	NBRP, Japón
IC5930	NECB1 *gidA::Tn10* (clone 4)	Bregeon, 2001
IC5931	NECB1 *gidA::Tn10* (clone 8)	Bregeon, 2001
IC5932	NECB1 *gidA::Tn10* (clone 10)	Bregeon, 2001
IC5695	IC4639 *gidB::Kan*	This study
IC5936	IC4639 *gidA::Kan*	This study
IC5933	IC4639 *gidA::Tn10* (IC5930 mutation)	This study
IC5934	IC4639 *gidA::Tn10* (IC5931 mutation)	This study
IC5935	IC4639 *gidA::Tn10* (IC5932 mutation)	This study
IC5956	TOP10 + pIC1343	This study
IC5959	IC5936 + pIC1345	This study
IC5960	IC5695 + pIC1345	This study
	Plasmids	
pIC1343	pBAD-TOPO *gidB*-Flag	This study
pIC1345	pBAD-TOPO *gidA-gidB*	This study
pIC552	Transcriptional system for *lacZ* fusions	Macian, 1994
pIC1344	pIC552 + fragment 1,544 to 1,953 of *gid* operon (operon starts at the *gidA* ATG)	This study
pIC1371	pIC552 + fragment 1,544 to 1,890 of *gid* operon	This study
pIC1372	pIC552 + fragment 1,739 to 1,953 of *gid* operon	This study

pIC1373	pIC552 + fragment 1,739 to 1,890 of *gid* operon	This study
pIC1374	pIC552 + fragment -162 to -1 of *gid* operon	This study
pIC1460	pIC552 + fragment 1,838 to 1,890 of *gid* operon	This study
pIC1461	pIC552 + fragment 1,739 to 1,842 of *gid* operon	This study

ANTIBODY PRODUCTION

The polyclonal antibodies against GidA had been previously obtained by our group (Yim et al., 2006). Rabbit antisera for GidB detection was obtained by cloning *gidB* fused to the Flag sequence in the pBAD-TOPO vector (pIC1343). The GidB-Flag recombinant protein was over-expressed in TOP10 cells by induction with 0.05% L-Arabinose at 37ºC for 3 hours with moderate and permanent shaking. After induction, cells were recovered by centrifugation at 3000 g and washed with TBS (NaCl 150 mM; Tris 50 mM; pH7.5). Finally, cells were resuspended in TBS and lysed by short and repeated ultrasound pulses. The soluble extract was recovered by centrifugation at 16000 g for 30 minutes at 4ºC. The extract was incubated with ANTI-FLAG M2 Affinity Gel (Sigma) resin according to the manufacturer's instructions. After 1 hour of incubation at 4ºC with permanent, gentle shaking, the flow-through was discarded and the Anti-Flag resin was washed eight times with 10 mL of TBS + 0.01% Triton X-100. Recovery of GidB-Flag was performed with a seven-time elution of Glycine 0.1 M pH 3.5 (one volume of resin per elution). Immediately after the elution, GidB-Flag was buffered in Tris-HCl 1M pH 7.5. The eluted protein was washed and concentrated in AMICON ULTRA Ultracel-10k filters (Millipore). The GidB-Flag extract was analysed by SDS-PAGE and was Coomassie Blue stained. Five inoculations (one inoculation every fortnight) containing 1mg of the GidB-Flag recombinant protein and Freund's adjunvant (Sigma) were injected in New Zealand rabbits. The anti-GidB activity of rabbit sera was evaluated one week after inoculation by Western blotting using a monoclonal Anti-Rabbit IgG-peroxidase (Sigma). Finally, the immuno-purification of the polyclonal antibodies was achieved by the fixation of up to 2 mg of GidB-Flag on a nitrocellulose membrane followed by blocking and incubation with different antisera for 1 hour at room temperature. The membrane was washed twice with TBS + Igepal 0.1% (Sigma). The elution of the specific GidB-Flag antibodies was done by incubation with 0.2 mL of glycine 0.1 M pH 3.5 for 5 minutes. After elution, glycine was buffered with 0.05 mL of Tris-HCl 1M pH7.5. The anti GidB-Flag activity in the eluent was evaluated by Western blotting as indicated below.

Western Blotting

Mutant strains IC5936, IC5933, IC5934, IC5935, IC5695, and parental IC4639 were cultured overnight a 37ºC in LB media supplemented with the respective antibiotics. The next day, all the cultures were diluted to 1/50 in 5 mL of LB media without antibiotics and cultured for 150 minutes at 37ºC. Cultures were maintained in a steady-state by diluting to OD_{600} = 0.2. Then, 4 mL of OD_{600} = 0.7 – 0.8 cultures were centrifuged at 3000 g and 4ºC. Cells were diluted in TBS and lysed by sonication. Soluble fractions were recovered by centrifugation at 16000g and 4ºC for 20 minutes. Protein concentration was measured through the Bradford assay (Bio-Rad Protein Assay) and the standard curve was titrated with BSA. Then 200 µg per protein extract were analysed in SDS-PAGE using the BenchMark Prestained Protein Ladder (Invitrogen). Denatured proteins were transferred to the nitrocellulose membrane, and overnight incubation with anti-GidA (1/5,000) and anti-GidB (1/5,000) was achieved after membrane blocking. The next day, membranes were washed twice with TBS + 0.1% Igepal. Incubation with Anti-Rabbit peroxidase (1/5,000) was done for 2 hours. Finally, the detection of the native Gid proteins was perfomed with ECL Western Blotting Detection Reagents (GE Healthcare) following the manufacturer's instructions.

Streptomycin Resistance Assay

Strains IC4639, IC5695, IC5935, and IC5936 were cultured in LB media overnight at 37ºC. Night cultures were diluted to 1/50 in LB media and incubated at 37ºC with permanent shaking until reaching the OD_{600} = 0.5 – 0.7 values. Then, cultures were spotted on LB agar supplemented with a minimum inhibitory concentration of streptomycin (20µg/mL) determined for the wild-type strain IC4639 in accordance with the previous method (Andrews, 2001).

In Silico Analysis of the *gid* Operon Sequence

A sequence between positions -378 and 2,577 of the *gid* operon (positions 3,921,080 to 3,924,034 of the NC_000913 GenBank entry) was submitted in the *Neural Network Promoter Prediction* (Reese, 2001) and *BPROM* (http://www.softberry.com/berry.phtml) web servers. Predictions from both servers were plotted in Figure 2A with their respective prediction values. String server (Snel et al., 2000b; von Mering et al., 2007) was used to evaluate the clustered fashion of *gidA* and *gidB* in the bacterial genomes. After this

analysis, the sequences of the respective *gid* operons were analysed with a view to seeing if they contained an inter-gene region by a direct inspection of the sequences in KEGG database (Kanehisa, 2002; Kanehisa et al., 2008). The inter-gene region of the *Escherichia coli gid* operon, comprised between positions 1,891 and 1,953 of the *gid* operon, was analysed by the *mfold* server to predict the secondary structure of mRNA (Zuker, 2003). This last procedure was also performed for those organisms in which an inter-gene sequence was found in the *gid* operon.

Upstream *gidB* Transcription Activity

Experimental testing of *in silico* predictions was carried out by cloning the predicted promoter region fused to the *lacZ* reporter (Macian et al., 1994). Initially, a 410 bp region (positions 1,544 to 1,953) was cloned in plasmid pIC552 using the respective primers with *XhoI* targets. To limit the localisation of the promoter region, different overlapping segments of the original 410 bp region were also cloned individually in the pIC552 plasmid. All the different clones carrying the respective constructed plasmids were evaluated and measured for the β-galactosidase activity produced *in vivo* (Miller, 1990). The deviation of the β-galactosidase activity derived from the plasmid copy number was normalised with β-lactamase activity (Andrup et al., 1988). The β-galactosidase activity derived from the *gidB* promoter was compared to the activity produced by the *gidA* promoter which was also cloned in the pIC552 plasmid as a control. The β-galactosidase activity derived from plasmids pIC1373 and pIC1374 during cell growth were measured as described above. Plasmids pIC1460 and pIC1461 carrying small fragments (53 bp and 104 bp, respectively) of that cloned in pIC1373 were equally treated and compared with the last type of activity produced.

Half-Life Measuring

The IC5959 and IC5960 strains carrying the pIC1345 plasmid were cultured in LB media and 100 μg/mL Ampicillin (Apollo Scientific) at 37ºC. Overnight cultures were diluted to 1/100 in the same media supplemented with 0.05% L-Arabinose (Sigma). Cultures were incubated at 37ºC with permanent shaking for 2 hours. The expression of the *gid* operon was detained in both strains by briefly centrifuging cultures at 3000g for 10 minutes and with the subsequent dilution in the same volume of LB supplemented with 1% glucose. Cultures continued growing at 37ºC and culture fractions were recovered at 0, 15, 30, 60, 90, 120, 150, and 180

minutes after the addition of glucose. Fractions were stored on ice until processed by Western blotting. Soluble fractions and further analyses were carried out according to the previously described methods (see Western Blotting procedures). Band densitometry was assisted by the *ImageQuant TL v2005* (Amersham) software. The signal was normalised with the total amount of proteins extracted per fraction. Fitting the exponential decay and half-life determination were calculated by the *GraphPad Prism v4.0*.

Results

Expression profile of GidA and GidB in mutants

Our initial intention was to study the expression pattern of the Gid proteins in a set of different mutants carrying insertion elements in either *gidA* or *gidB* (Figure 1A). Using steady-state cultures in the log phase (OD_{600} = 0.6 – 0.8), we analysed the presence of native GidA and GidB in soluble cell extracts (Figure 1B). As expected, no protein corresponding to GidA was detected in the different *gidA* mutants. In these mutants, however, the detection of the GidB product was always possible, at least in low proportions (see mutants IC5933, IC5934 and IC5935). Unlike the *gidA::Tn10* mutants, GidB accumulated in the *gidA::kan* mutant (IC5936) at levels similar to those observed in wild-type cells.

Besides, we also decided to study the phenotypic effect of the attenuated GidB expression in the *gidA::Tn10* mutants. *gidB* mutants are known for their low-level streptomycin resistance (Okamoto et al., 2007). Therefore, we tested the ability of the *gidA* mutants to grow in the presence of LB with 20μg/mL of streptomycin (Figure 1C). After an overnight incubation at 37ºC, the *gidA* mutants showed no streptomycin resistance as evidenced in the *gidB* mutant IC5695. Therefore, despite the lower expression of *gidB* seen in all the *gidA::Tn10* mutants, it is apparently sufficient to efficiently modify most of rRNA 16S and to avoid the expected streptomycin resistance associated with a lack of m^7G527 modification (Okamoto et al., 2007).

The expression of GidB observed in the *gidA* mutants could be due to: i) the presence of a specific promoter of *gidB* in *gidA* ORF; ii) the activity of a Tn10 promoter or; iii) the transcription from the *gidA* promoter (*gidAp*) that is able to overcome the Tn10 polar effect.

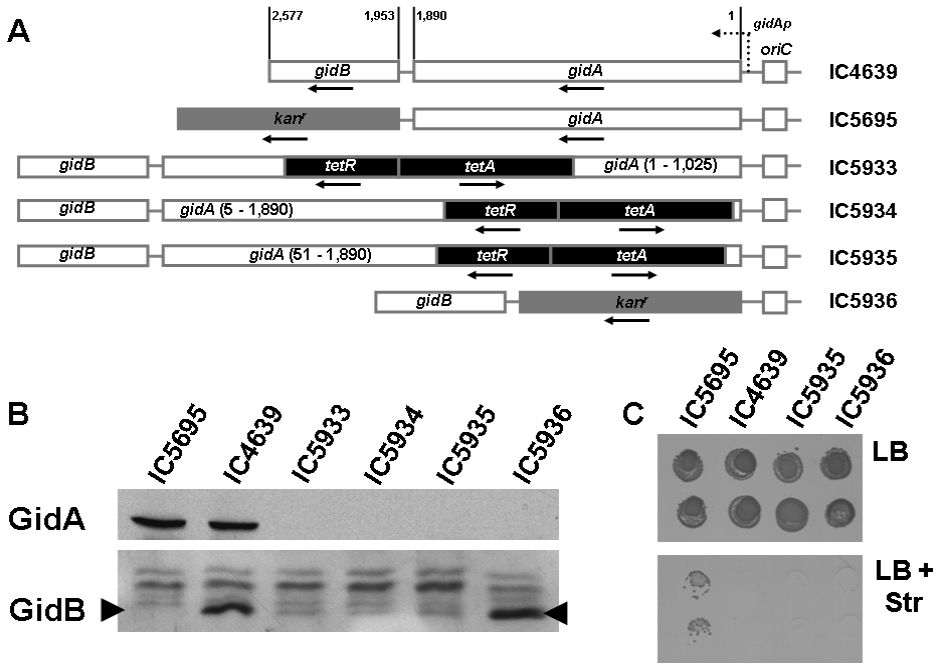

Figure1. The profile expression of the Gid proteins in mutants. A – Mapping of the *gid* operon in all the mutant strains used to study the Gid proteins expression. Black arrows show the reading orientation of the *gid* genes and the elements inserted into the *E. coli* chromosome for the respective mutants. B – A Western blot assay showing the expression profile of the Gid proteins in all the different strains drawn in Figure 1A. C – Streptomycin resistance assay to test the GidB function in the *gidA* mutants.

If the GidB expression in the *gidA:Tn10* mutants were due to the activity of a *gidB* promoter, such a regulatory element may be located downstream of the Tn10 insertion site. A comparison of the Tn10 insertion site in the *gidA:Tn10* mutants defined the region between positions 1,025 and 1,953 of the *gid* operon as being the most probable region to contain the *gidB* promoter (see Figure 1A). We decided to test this hypothesis given its relevance to determine further expression features of the *gid* operon.

PREDICTION AND CONFIRMATION OF THE SPECIFIC PROMOTER FOR *GIDB*

To test the presence of additional regulator elements in the *gid* operon, we initially performed an *in silico* analysis using the entire sequence of *gidA* and the *gidB* genes of *Escherichia coli K12,* including the 63 bp inter-gene region.

In addition, a segment of the *oriC* region was included in the analysis as a control to predict the *gidA* promoter. Promoter prediction was assessed by two different algorithms (see Methods); their respective predictions are shown in Figure 2A. In both cases, sequence-based algorithms promoters were used to predict prokaryote signals. In addition to the respective scoring values per promoter prediction, we focused on the consensus localisation of those predictions generated in both methods. Thus, both predictors agreed to detect the promoter signals in two specific regions along the operon. The first corresponded to positions -92 to -115 where *gidAp* is localised according to a previous study (Kolling et al., 1988). The location of the second region was predicted in the 3' region of *gidA*, between positions 1,881 and 1,885. This latter prediction is coherent with our hypothesis for a specific *gidB* promoter (*gidBp* from this point onwards) in this region. Interestingly, the putative *gidB* promoter is upstream from an inverted repeated region mapped in the region between *gidA* and *gidB* (Walker et al., 1984). This inverted repetition is able to generate a stable stem-loop secondary structure (Figure 2B) according to RNA fold predictions. When this region was analysed for sequence homology in other bacterial organisms, a poor sequence similarity was found (except in the *Enterobacteriaceae* family). However, a similar distribution of this secondary structure was found in the following bacteria families: *Enterobacteriaceae*, *Vibrionaceae*, and *Shewanellaceae* belonging to Gammaproteobacteria; and *Clostridiaceae* family belonging to Firmicutes. Because of the poor sequence similarity of the stem-loop among the different organisms analysed, it is likely that this operon feature is the result of evolutive convergence to regulate the expression pattern of the *gid* operon in accordance with growth requirements.

In the species in which this region was found, it was present with a variable sequence extension (63 to 180 bp, approximately). Stem-loop structures in mRNA are frequently associated with transcription terminator activity because they are able to destabilise either intrinsically or in a Rho-mediated manner: the RNA Polymerase complex during elongation. According to canonical signals of Rho-independent terminators (Bachellier et al., 1996), the stem-loop located between *gidA* and *gidB* cannot be classified as such.

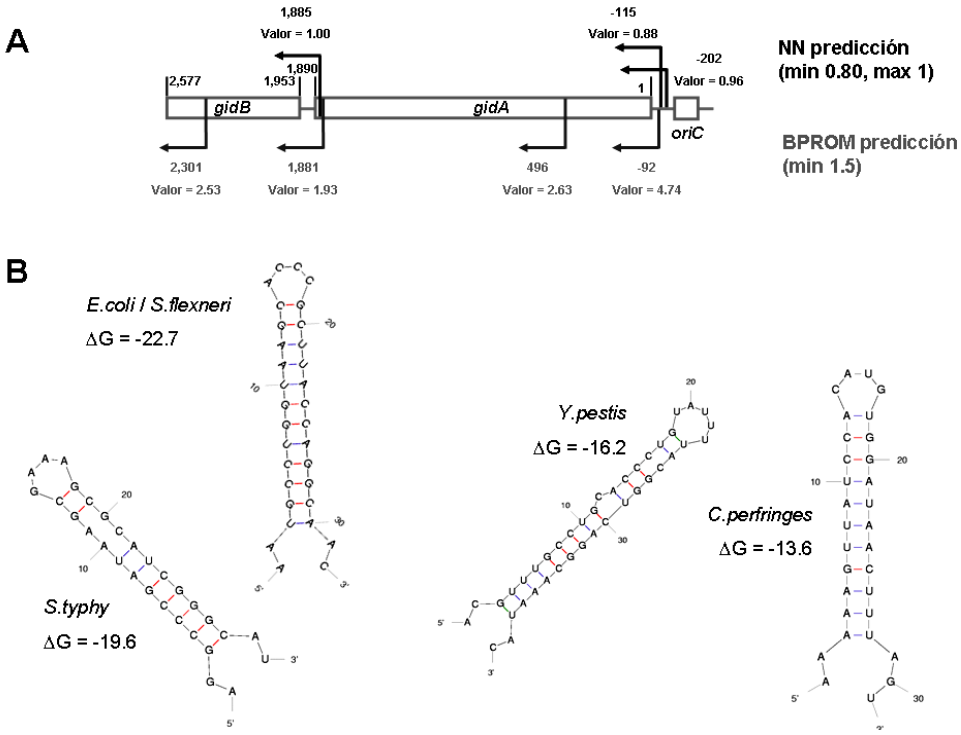

Figure 2. The *in silico* prediction of the new regulatory elements in the *gid* operon. A – Map of the *gid* operon showing promoter predictions of two bioinformatic algorithms. The promoter predictions done by the *Neural Network Promoter Prediction* server (Reese, 2001) are drawn at the top of the *gid* map with their respective prediction values. Those predictions done by the BPROM server are depicted at the bottom of the *gid* map with the respective predictions values. B – Stable secondary structures predicted from the inter-gene *gidA-gidB* regions of the different organisms where *gid* genes were not overlapped. Species and the free energy calculated by the mfold server (Zuker, 2003) per structure are also shown.

Several DNA fragments containing the putative regulatory signals were cloned in the *lacZ* reporter vector (see Methods). As shown in Figure 3, the transcriptional activity generated by the 1739-1890 region (pIC1373) is abated when the region comprised between positions 1890 and 1953 is also included in the insert (pIC1372). The fact that the transcriptional activity measured in plasmids pIC1372 and pIC1373 is due to the predicted regulatory signals, but not to cloning artefacts, is corroborated by plasmids pIC1344 and pIC1371 where the DNA fragments differing from those carried by pIC1372 and pIC1373 have been cloned. Therefore, we conclude that the *gidB* promoter predicted by *in silico* methods has activity *in vivo* and that the

stem-loop works as a factor-dependent transcriptional terminator. We predict that this terminator also abates the transcription mediated by the *gidA* promoter given that it resides in a non-traslated region.

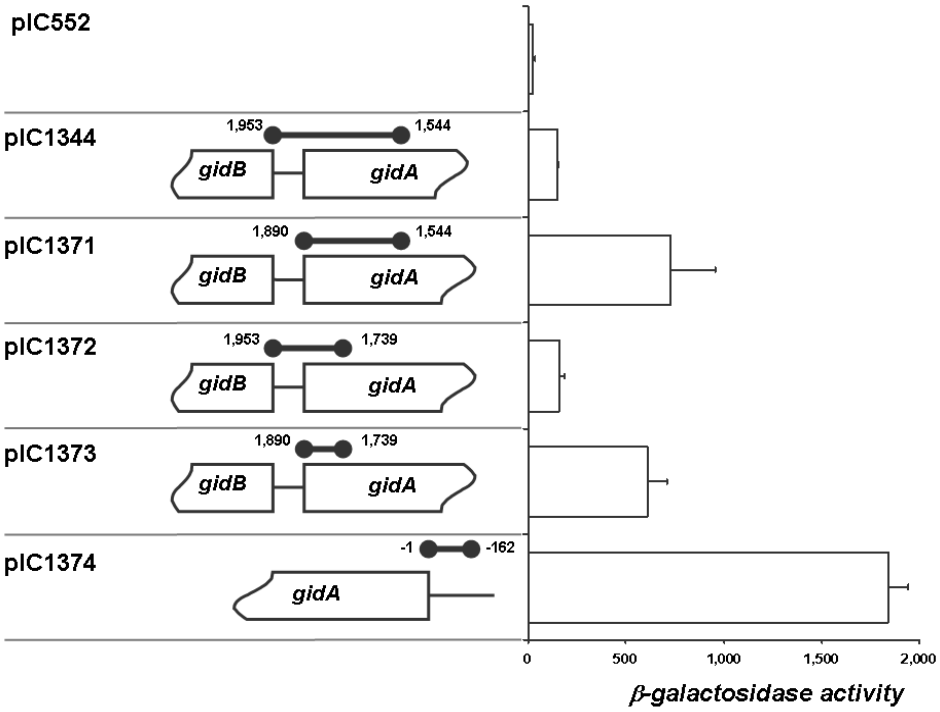

Figure 3. The transcriptional activity derived from the predicted regions acting as the new regulatory elements of the *gid* operon. pIC552 and derived plasmids containing different segments of the *gid* operon were tested for β-galactosidase activity. To the left, the plasmids containing limited and numbered (see the bars and positions of the *gid* operon per plasmid) regions of the *gid* operon are shown. To the right, the β-galactosidase activity generated by the respective plasmids is plotted. Measurements and standard errors were calculated from three different replicates.

The intrinsic activity of *gidBp* corresponds to 0.30-0.50-fold *gidAp* activity (by comparing the activities derived from vectors pIC1373 and pIC1374). In addition to the activity found in *gidBp*, we detected a transcription termination activity when the short segment of 63 bp between *gidA* and *gidB* ORF (this segment is called *gidAt* from this point onwards) was included in *lacZ* fusions (see the activities of plasmids pIC1344 and pIC1372 in Figure 3). Thus, we confirm the transcription terminator role of the stem-loop structure present in this region. This terminator reduced the

transcriptional activity of *gidBp* by almost 4-fold, and we assume that *gidAp* activity needs to be also decreased in the same manner.

By splitting the shortest region of ≈152bp cloned in pIC552 (pIC1373), we managed to outline the specific DNA region of *gidBp.* According to *in silico* predictions, we cloned a 53 bp region between positions 1,838 and 1,890 where boxes -35 and -10 were predicted to appear (pIC1460), whereas the remaining 104 bp region between positions 1,739 and 1,842 was used as a negative control (pIC1461). After measuring the β-galactosidase activity produced by these new pIC552-derived plasmids, we confirmed the presence of the main regulatory boxes in the short 53 bp region cloned in pIC1460 (Figure 4) whose β-galactosidase activity did not significantly differ from that generated from pIC1373 ($p<0.2694$). In contrast, it was ruled out that the remaining 104 bp region does not promote transcription activity ($p<0.0353$). The most probable sequences corresponding to boxes -35 and -10 are depicted in Figure 4b, according to the *in silico* predictions.

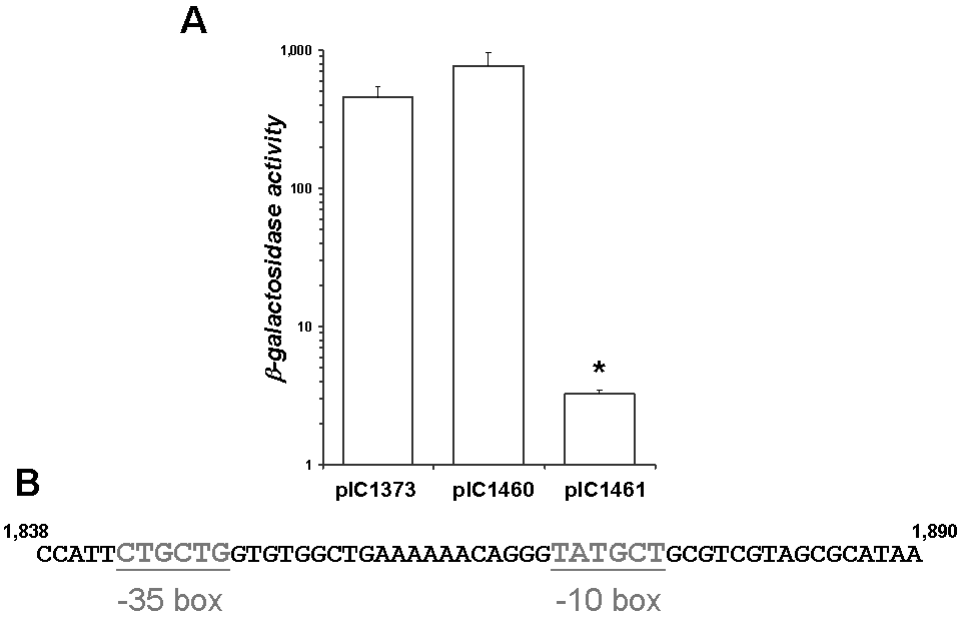

Figure 4. Dissection of the specific region acting a *gidBp*. A – Transcriptional activity from two additional pIC552-derived plasmids. A 152bp region was divided into vectors pIC1460 and pIC1461. pIC1460 contains a very short 53 bp region where regulatory elements such as boxes -35 and -10 were predicted to localise, while pIC1461 contained the remaining 104bp DNA region used as a negative control of transcriptional activity. B – The 53nt DNA sequence cloned in pIC1460 with its predicted regulatory elements.

Expression of the *gidA* and *gidB* promoters throughout cell growth

Given that the expression at *gidB* depends on two different promoters, we decided to study their expression pattern during cell growth and thus evaluate a possible role to modulate the expression of *gidB* under different growth conditions (Figure 5). Previous analyses have shown that *gidAp* is strongly regulated during cell division. Consequently, *gidAp* activity decreases after the initiation of chromosome replication (Theisen et al., 1993; Ogawa & Okazaki, 1994). Our results show that the *gidAp* expression pattern decreases as the cell culture grows. The highest *gidAp* activity was found very early on in the log phase, while the lowest activity was seen immediately prior to entry in the stationary phase. Finally, *gidAp* activity slightly increased during the stationary phase. This last feature could be the result of some DNA elements present in *gidAp* that could be both recognised and regulated by the proper σ factors of the stationary phase as σ^{38}. To test this possible effect of σ^{38}, we measured the β-galactosidase activity generated from plasmid pIC1374 in a wild-type strain, MC4100, and its *rpoS*::Tn10 *E coli* mutant (OD_{600} = 1.2 to 1.3), both kindly donated by Dr. Miguel Vicente (Centro Nacional de Biotecnología – CNB, Madrid, Spain). The results show that the β-galactosidase activity promoted by *gidAp* did not significantly differ in both strains (data not shown). Therefore, the slight increase in the *gidAp* transcriptional activity in the stationary phase of cells was σ^{38}-independent.

On the other hand, *gidBp* showed a lower expression pattern than *gidAp* which slightly increased from OD_{600} = 0.5 to manage a 2.5-fold-increase of its initial expression in the stationary phase. The extra contribution of *lacZ* mRNA granted by *gidBp* could indicate that *gidBp* activation compensates the decreasing *gidAp* activity in the log phase.

Half-life of the GidA and GidB proteins

Recently, titration of the GidA and GidB proteins was achieved by reliable methods like Mass Spectrometry (Ishihama et al., 2008). GidB titration in the log phase (OD_{600} = ≈0.4) was 268 molecules per cell, while it was 212 for GidA. Despite the novel regulatory signals decribed herein, at the DNA level, the global balance of the transcriptional and translational regulation for the expression of *gidA* and *gidB* show no significative differences between the GidA and GidB levels. To investigate this parameter, we studied the *in vivo*

stability of the Gid proteins. Using a controlled expression of the *gid* operon (under the control of P_{BAD}) inducible by L-arabinose and repressed by glucose in the media, we were able to measure the half-life of these proteins (Figure 6).

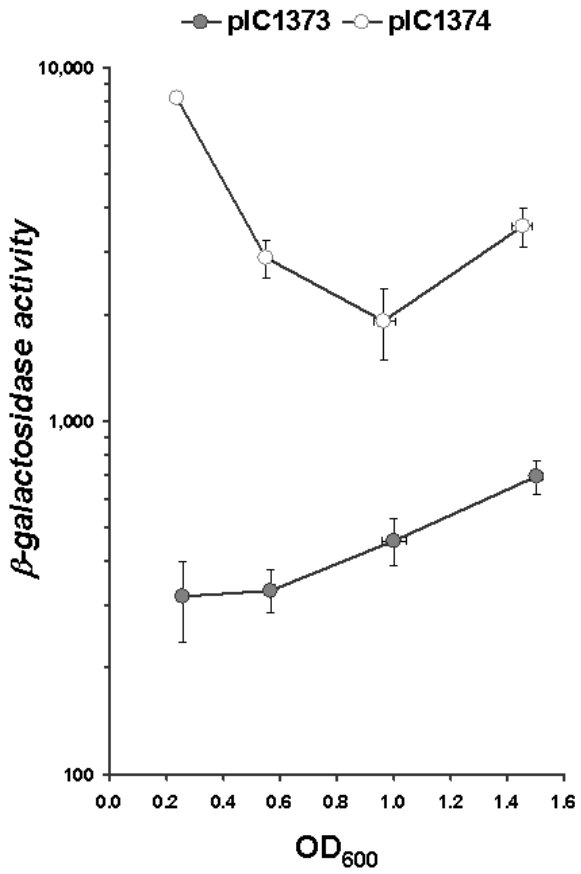

Figure 5. The transcriptional activity of promoters *gidAp* and *gidBp* during cell growth. A - Plasmids pIC1373 and pIC1374 containing the DNA regions of *gidBp* and *gidAp,* respectively, were used to measure the β-galactosidase activity derived from them at the different stages of cell growth. Measurements and standard errors were obtained from three different replicates.

After fitting the exponential decay for the respective proteins, the half-life of GidA was determined in 31.4 ± 3.5 minutes whereas the half-life of GidB was calculated in 16.9 ± 1.3 minutes (Figure 6). Thus, GidA shows 1.9-fold more stability than GidB.

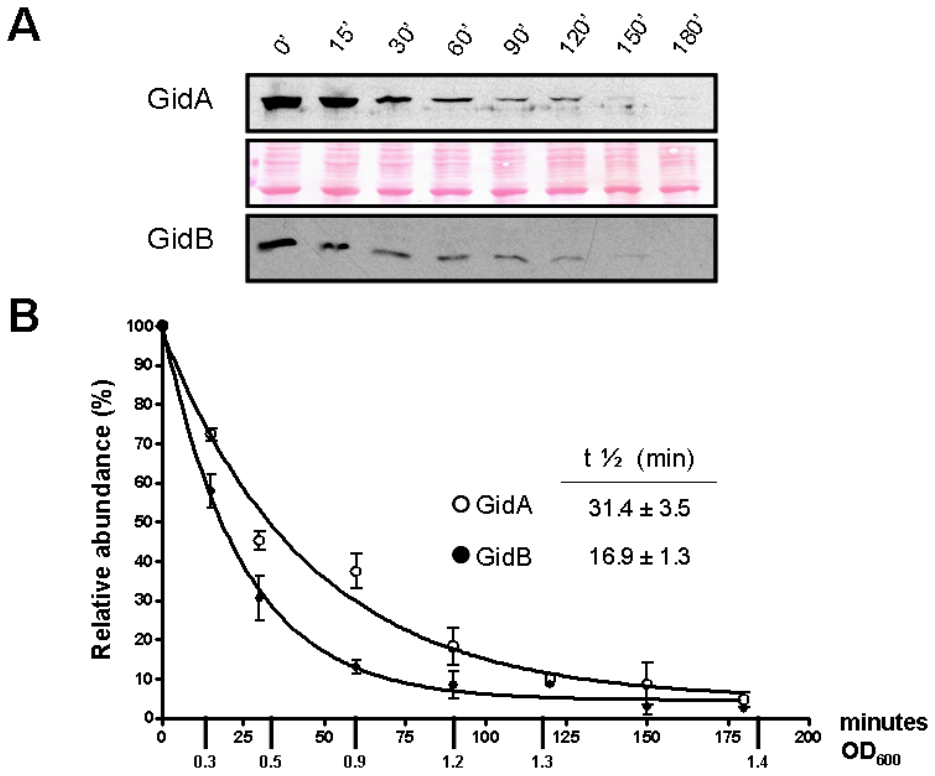

Figure 6. Determining the half-life of the Gid proteins. A – Western blot showing the respective degradation patterns for GidA and GidB at time intervals of 0, 15, 30, 60, 90, 120, 150 and 180 minutes after detaining the induction of the *gid* operon. After the western blot had finalised, a nitrocellulose membrane was stained with Red Ponceau and presented as a protein loading control. B – The data from the kinetic degradation of proteins GidA and GidB were plotted. Half-life values were calculated from three different experiments during culture growth.

Discussion

In agreement with previous results, we have shown that the expression of *gidA* and *gidB* derive mainly from *gidAp*. As a result, when *gidA* is interrupted by exogenous DNA elements, the *gidB* expression attenuates, but does not completely cease. This effect is only seen when Tn10 is used as an insertion module. Conversely, interruption with the *kan*^{r} element does not affect the expression of *gidB*. Additionally in a few cases, turn-on polarity has been described to allow the expression of adjacent and distal DNA regions caused by the internal promoters of insertion elements (Ciampi et al., 1982). Although the expression of *gidB* seen in the *gidA::Tn10* mutants may be a result of the promoters in Tn10, we have shown that it depends on the *gidBp*

identified in this work by bioinformatic approaches. All the mutants used in our study were inactivated with mini-transposons (which are known not to produce polar effects (Kleckner et al., 1991)). Therefore, the native GidB expression seen in the IC5936 mutant (*gidA::kan*) is not likely to be the result of the turn-on polar effects of the *kan*^{r} element.

The *in silico* predictions performed in this study to discern additional regulatory features in the *gid* operon were successfully corroborated by experimental approaches. This enabled us to report a specific *gidBp* promoter localised between positions 1,739 and 1,890 of the *gid* operon. A more specific localisation of the *gidBp* region determined that only a fragment between 1,838 and 1,890 cloned in pIC1460 is necessary to promote transcription, even at the higher yet not significantly different levels generated from pIC1373. Likewise, the transcription terminator *gidAt* was found within 63bp which separates the *gidA* and *gidB* genes in *E. coli* (see Fig. 2). Under our experimental conditions, *gidAt* decreases the transcriptional activity generated by *gidBp* by 75% (Fig. 3). This terminator does not exhibit the canonical signals of intrinsic transcription terminators such as C-rich content and poly-U ending (Bachellier et al., 1996). Therefore, its activity is possibly mediated by transcriptional factors, thus opening a way for the modulating expression of *gidB* in coordination with cellular metabolism.

The study of *gidAp* shows that its maximum activity takes place in very early stages of cell growth (Figure 5). This type of regulation resembles that known for rRNA and tRNA genes. Indeed, ribosomes and tRNA undergo strong regulation and their concentrations are 11-fold higher in fast-growing cells than in slow-growing cells (Bremer & Dennis, 1996). In line with this, the expression of the Gid proteins seems to fit their substrate production. *gidBp* activity does not directly correlate with that seen for *gidAp*. *gidBp* shows relatively low activity at early stages of cell growth, and its expression increases until the stationary phase. Although *gidAp* and *gidBp* appear to promote transcription in the stationary phase, we discard the involvement of σ^{38} in this activity (data not shown).

The additional expression of *gidB* caused by the activity of promoters *gidAp* and *gidBp* is supported by a global analysis of mRNA abundance in *E. coli* (Bernstein et al., 2002) (data available at http://chase.ou.edu/oubcf/)

where *gidB* shows a higher expression profile than *gidA*. This result could be due to the synergic activity of both the *gidAp* and *gidBp* promoters, conferring *gidB* a higher expression. Despite *gidB* being more expressed, the number of molecules between both Gid proteins is almost identical (Ishihama et al., 2008). Thus we realise that protein GidB degradates more rapidly than GidA. Therefore, our results are in agreement with the observations reported by Ishihama and co-workers.

Acknowledgements

We especially acknowledge the National BioResource Project (NIG, Japan) and its *E. coli* mutant strain collection (KEIO) for kindly donating mutants IC5695 and IC5936, and Dr. Bregeon for donating IC5930, IC5931, and IC5932 strains.

Chapter 2
Functional and Biochemical Insights of *E. coli* RsmG Methyltransferase

Summary

Methylations are predominantly present in the bacterial ribosome RNA. Modifications in rRNA 23S can be found in the different structural domains interacting with the tRNAs and the small ribosome subunit. However, modifications in rRNA 16S seem to have a more restricted localisation. Most of them converge in the decoding centre of the 30S subunit. Moreover, some are evolutionarily conserved, at least in the bacteria kingdom. One of these conserved modifications is m^7G527 which is found in rRNA 16S of *E. coli*. This methylation is accomplished by RsmG methyltransferase. S-adenosyl-L-methionine (AdoMet)-dependent methylases consist in the most representative set of proteins to acquire a common molecular role by evolutive convergence. Therefore, the study of the sequence-structure-function relationships in these proteins could shed

light on the evolution mechanisms acting on them. The aim of this study is to provide new functional insights of the m^7G527 modification present in rRNA 16S of *E. coli*. Similarly, we evaluated the functional and biochemical aspects of its responsible enzyme, RsmG, through directed mutagenesis. We found that a lack of m^7G527 affects the read-through rate of the UAG stop codon and that it is also involved directly in reading frame maintenance. Besides, we describe a set of residues where AdoMet, and probably rRNA binding functions, lie.

INTRODUCTION.

Post-transcriptional modification is a common feature in non-coding RNA such as tRNA or rRNA. Modifications on bacterial tRNAs have been clearly associated with decoding functions for a proper codon·anticodon interaction (Elseviers et al., 1984; Bregeon et al., 2001; Nasvall et al., 2007). They not only support the codon sensitivity and codon choice, but are also involved in tRNA stability, tRNA folding and in maintaining the well-structured anticodon domain to decode mRNA properly (Björk & Hagervall, 2005a; Agris, 2008). Modifications in bacterial rRNA are less abundant than tRNA (approximately 1 per hundred of ribonucleotides), and their function in mRNA decoding is not clear. The most common bacterial rRNA modification is the addition of the methyl group to the purine/pyrimidine rings or in position 2' of the ribose moiety of ribonucleotides. In *Escherichia coli,* the full set of methylations consists in ten modifications in rRNA 16S, while 15 can be found in rRNA 23S (Ofengand & Campo, 2005a). At the same time, most methylations have been associated with the respective enzymes responsible for them. As a result, current orphan methylation in rRNA 16S in *E.coli* is m^2G1516, and m^6A2030, m^7G2069, Cm2498 and mdC2501 in rRNA 23S (briefly reviewed in (Purta et al., 2009)).

RsmG is one of methyltransferases that modifies rRNA 16S in *Escherichia coli.* Although early studies associated *rsmG* (known as *gidB*) with the phenotype of *glucose-inhibited cell division* in *E. coli*, this association was indirectly based on its cluster localisation with *mnmG* (*gidA*) close to the chromosome replication origin *oriC*. (von Meyenburg et al., 1982).

RsmG (*gidB*) encodes a 207 aa in the length protein in *E.coli*. The phylogenetic distribution of RsmG homologues is restricted to bacteria and

plants. The *E. coli* RsmG *x*-ray structure revealed that this protein has a classical methyltransferase fold (Romanowski et al., 2002). Nowadays, it is well-known that RsmG produces the S-adenosyl-L-methionine (SAM)-dependent m^7G modification in position 527 of rRNA 16S in *Escherichia coli* (Okamoto et al., 2007). Given the vast distribution of *rsmG* in bacterial genomes, the meaning of m^7G modification in rRNA 16S could be critical for decoding ribosome properties (von Mering et al., 2007). However, neither growth effects nor UAG read-through have been described for *rsmG* mutants (Okamoto et al., 2007). Notwithstanding, *rsmG* mutants show low-level streptomycin resistance (Nishimura et al., 2007a; Nishimura et al., 2007b; Okamoto et al., 2007). On the other hand, *rsmG* mutants have an emerging frequency of high-level streptomycin about 200-fold higher than wild-type ones (Okamoto et al., 2007). Although the mechanism producing high streptomycin resistance remains unknown, most of these streptomycin resistants show mutations in the *rpsL* gene that encodes protein S12 of the 30S ribosome subunit.

No relevant information about the sequence-structure-function relationships of RsmG is available. Based on sequence and structural *in silico* analyses with experimental corroboration based on directed mutagenesis, we have determined some RsmG residues that prove critical to accomplish the m^7G527 modification in rRNA 16S of *Escherichia coli*. The results obtained in this RsmG study will be useful for exploring methylation mechanisms in other rRNA methylases, most of which have been completely unexplored in this context.

Methods

Sequence and Comparative Structural analyses

The phylogenetic distribution of RsmG was studied with the STRING server (von Mering et al., 2007). More than 30 RsmG sequences belonging to the organisms of the most representative bacteria groups were aligned using the MUSCLE application and default parameters (Edgar, 2004). A Hidden Markov Models-based profile was built using the HMMER v. 2.3.2 application and default parameters (El Yacoubi et al., 2009). Highly conserved residues of RsmG proteins revealed by HMM-profile were visualised through the Logomat-P server (Schuster-Bockler & Bateman,

2005). The HMM profile of the RsmG family was used as a bait in a Blastp search to find close homologues in the PDB database, where S-adenosyl-L-methionine was a co-factor in the structure. Manual docking of AdoMet was done by superimposing the alpha carbons of glycines, thus conforming the canonical SAM-binding motif G-X-G-X-G using a DeepView viewer (Guex & Peitsch, 1997). Molecular graphics images were produced with the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (Pettersen et al., 2004).

Strains and plasmids

The BW25113 *rsmG::kan* strain was kindly donated by the National BioResource Project (NIG, Japan) along with its *E. coli* mutant strain collection (KEIO). Using the P1 transduction system (Miller, 1990), the *rsmG* mutation was recovered in the Dev16 background to study the read-through phenotype. The right insertion of the *rsmG::kan* mutation was confirmed by PCR using the primers hybridising in the flanking DNA sequences of *rsmG* and internal primers for kan^r gene amplification (Datsenko & Wanner, 2000). *E. coli rsmG* was amplified from the DNA obtained from MC1000 and cloned in a pBAD-TOPO vector with a FLAG sequence at 3'. The full set of oligonucleotides used for site-directed mutagenesis is shown in Table 1. To study AdoMet affinity, the wild-type and mutants obtained in pBAD-TOPO were subcloned in the pET15b vectors using primers with *NcoI* and *BamHI* sites (see Table 1).

Read-through Assay

Misreading the UAG stop codon carried by the *lacZ105* gene was determined by using the β-gal assay, as described by Miller (Miller, 1990). Additionally, overnight and refreshed cultures were supplemented with 1mM IPTG to increase the β-galactosidase activity. Comparison of mutant and wild-type activity was made for statistical significance, which was determined by five independent replicates analysed by a t-test and Welch's correction.

Frameshift Assay

Effects of the *ΔrsmG* mutation on the programmed *dnaX* frameshifting were investigated by transforming mutant and wild-type cells with a pBAD22 plasmid carrying a full-length wild-type version of the recombinant

dnaX E.coli gene tagged at the 5' end with 6His. The ratio of the γ and τ isoforms was measured by band densitometry of anti-His western blotting in both wild-type and *ΔrsmG* strains. Statistical significance was determined by four independent replicates and was analysed by a *t-test* using Welch's correction. The +2 frameshifting was also studied for the *rsmG* mutant by transducing the *rsmG* null mutation to the NECB1 background, as previously reported for this assay (Bregeon et al., 2001).

Table 1. List of oligonucleotides used to produce all the mutants studied.

Oligonucleotides 5' – 3'	
Cloning	
pBAD-TOPO/*rsmG-flag*	ATACCATGGTGCTCAACAAACTCTCCTTACTGC
	TTATTA**TTTGTCGTCGTCGTCTTTATAGTC**AATTTTATTTGCTTTA ATCACCACC
pET15b/*rsmG-6His*	ATACCATGGTGCTCAACAAACTCTCCTTACTGC
	AGATAGGATCCTTATTA**ATGATGGTGATGATGGTG**AATTTTATTT GCTTTAATCACCACC
Mutagenesis	
N39A	ACAAAGCGTAC**GCC**CTGACTTCGGT
	ACCGAAGTCAG**GGC**GTACGCTTTGT
H53A	AGATGCTGGTACGC**GCT**ATTCTCGATAGCAT
	ATGCTATCGAGAAT**AGC**GCGTACCAGCATCT
D56A	CGCCATATTCTC**GCT**AGCATTGTGGTG
	CACCACAATGCT**AGC**GAGAATATGGCG
D71A	GTGAACGGTTTATC**GCT**GTCGGCACCGGACC
	GGTCCGGTGCCGAC**AGC**GATAAACCGTTCAC
G73A	GGTTTATCGATGTC**GCC**ACCGGACCAGGACT
	AGTCCTGGTCCGGT**GGC**GACATCGATAAACC
G75A	TCGATGTCGGCACC**GCA**CCAGGACTGCCAGG
	CCTGGCAGTCCTGG**TGC**GGTGCCGACATCGA
G77A	TCGGCACCGGACCA**GCA**CTGCCAGGCATTCC
	GGAATGCCTGGCAG**TGC**TGGTCCGGTGCCGA
P79A	CCGGACCAGGACTG**GCA**GGCATTCCACTCTC
	GAGAGTGGAATGCC**TGC**CAGTCCTGGTCCGG
D96A	ATTTCACTCTGTTG**GCT**AGCCTTGGTAAACG
	CGTTTACCAAGGCT**AGC**CAACAGAGTGAAAT
K100A/R101A	TGGATAGCCTTGGT**GCGGCC**GTGCGTTTCCTTCG
	CGAAGGAAACGCAC**GGCCGC**ACCAAGGCTATCCA
R123A	CCAGTACAGAGC**GCT**GTAGAAGAGTTT
	AAACTCTTCTAC**AGC**GCTCTGTACTGG
R139A	ATGGCGTAATTAGC**GCC**GCTTTTGCCTCTCT
	AGAGAGGCAAAAGC**GGC**GCTAATTACGCCAT
R139K	GCGTAATTAGC**AAA**GCTTTTGCCTC
	GAGGCAAAAGC**TTT**GCTAATTACGC
K165A	CTACGCGCTG**GCC**GGGCAAATGC
	GCATTTGCCC**GGC**CAGCGCGTAG

R197A	GGATGGCGAA**GCT**CATCTGGTGG
	CCACCAGATG**AGC**TTCGCCATCC

Expression levels of the recombinant RsmG protein

ΔrsmG carrying the pBAD-TOPO (Invitrogen) vector with the recombinant RsmG or mutations (with Flag tag at C-terminal) were cultured overnight in LB media supplemented with 100μg/mL Ampicillin (Apollo Scientific) at 37ºC. Overnight cultures were diluted to 1/100 in the same LB + Ampicillin and incubated at 37ºC with moderate and permanent shaking for 2.5 hours (OD_{600} = 0.6-0.7). Then, 2 mL of culture were centrifuged at 3000g for 10 minutes at 4 ºC. Cell pellets were resuspended in PBS pH 7.5 and lysed by short, repeated ultrasound pulses. Soluble fractions were recovered by centrifugation at 16000g for 20 minutes at 4ºC. Next, 300 μg of crude extract were analysed by Western blotting using anti-RsmG rabbit policlonal antibodies (previously obtained by our group, see Chapter 1) and anti-rabbit peroxidise (Sigma). The level of the expression of recombinant RsmG proteins was normalised against the expression of native RsmG observed in the wild-type strain carrying an empty pBAD-TOPO plasmid.

Streptomycin Resistance Assay

Cell cultures were recovered in the same way as described above. Subsequently, a small aliquot of each culture was set at OD_{600} = 0.4. Then, triplicate spots of the cultures diluted to 1/10 were plated on LB agar supplemented with 20 μg/mL of Streptomycin. Cell growth was evaluated after a 12-16-hour incubation at 37ºC.

In vivo complementation of the 16S rRNA modification

The *in vivo* complementation of the modification accomplished by RsmG in 16S rRNA was achieved using the same culture conditions described to study the expression levels of recombinant proteins. First, 60 mL of culture were processed to obtain a purified fraction of rRNA 16S. Cells were harvested by centrifugation and pellets were resuspended in 0.4 mL Buffer-A (25 mM Tris pH 7.4, 60 mM KCl, 10 mM $MgCl_2$). Next 2 mg of Lysozyme (Sigma) were added and the suspension was incubated at 37ºC for 20 minutes. Cell suspension was lysed by three-time freeze-thaw using liquid nitrogen. Then 0.6 mL of Buffer-B (Buffer-A supplied with 0.6% Brij35, 0.2% Na-deoxycholate, 0.02% SDS) and 0.1 mL of Phenol (citrate equilibrated, pH 4.3) were added and vigorously mixed. Suspension was incubated on ice for 15

minutes and the aqueous phase was twice extracted with phenol 1:1. RNA was precipitated overnight by adding 2.5 volumes of cold 99.5% ethanol containing 1% (w/v) potassium acetate. After washing the pellet twice with 70% ethanol, the pellet was dissolved in 2 mL buffer R200 (100 mM Tris-H_3PO_4 at pH 6.3, 15% ethanol, 200 mM KCl) to be applied to a Nucleobond AX500 column (Macherey-Nagel), pre-equilibrated with 10 mL of R200. The column was washed once with 6 mL of R200 and once with 10 mL of R750 (R200 supplied with 750 mM KCl instead of 200 mM). rRNA 16S was eluted with 6 mL of R1000 (supplied with 1000 mM KCl). Then, rRNA was precipitated by 0.7 volumes isopropanol, washed twice with 70% ethanol and dissolved in water. Afterwards, 25ug of rRNA were hydrolysed by nuclease P1 (Sigma) and overnight incubation in water with 1mM $ZnSO_4$ was followed by treatment with *Escherichia coli* alkaline phosphatise (Sigma) at pH 8.3 for at least 2 hours. The hydrolysed product was analysed by HPLC using a Develosil 5µ RP-AQUEOUS C-30 column (Phenomenex®) with a gradient elution to obtain optimal separations of the nucleosides. Buffer A had 2.5% methanol, 10mM $NH_4H_2PO_4$, pH 5.3, while buffer B had 25% methanol, 10mM $NH_4H_2PO_4$, pH 5.1. All the HPLC-nucleoside mutant profiles were compared with those derived from the wild-type ones. Normalisation and RsmG activity quantification per mutant were determined in peak area terms, as follows: $((m^7G/m^2G)_{mutant} - (m^7G/m^2G)_{null}) / (m^7G/m^2G)_{wild\text{-}type}$.

S-adenosyl-L-methionine Binding

All the RsmG mutants tested by *in vivo* complementation were cloned in pET15b in order to over-express the recombinant RsmG-6His proteins. AdoMet affinities were determined through Surface Plasmon Resonance (Biacore T100) by linking monoclonal anti-His immunoglobulines to CM5 chip using the Amine Coupling Kit (Biacore). The best conditions for protein immobilisation were achieved using PBS 0.25X pH 6.0. Then, 7µg of protein was immobilised for 150 seconds with flux at 10 µL/min. Different concentrations of AdoMet (New England Biolabs) were tested (ranging from 100 nM to 10 µM). The contact time was set at 40 seconds at the same flux as before. AdoMet affinities were calculated using the Biacore T100 Evaluation Software, V2.0 (Biacore).

Results

Read-through and frameshifting phenotypes of the *rsmG* mutant

Previously, the *Escherichia coli rsmG* mutant was reported to have no read-through activity other than observed in the wild-type Dev16 strain carrying the *lacZ*105(Amb) allele (Okamoto et al., 2007). Using a similar approach, we present the data where a mutation in *rsmG* causes a higher read-through rate of the UAG stop codon than that observed for the wild type (Figure 1A). The difference between Okamoto's result and that reported herein can be explained by the inducers of the *lacZ* expression used in our protocols (see Methods). The effect of *rsmG* mutation on the different frameshifting assays was also evaluated. In this way, a decreasing rate of the programmed frameshifting of *dnaX* ($p<0.0175$) was observed in *ΔrsmG* given the higher production of the τ isoform against γ (Figure 1B). No effect of the *ΔrsmG* mutation was observed on the +2 frameshifting assay (data not shown).

Sequence-structure-function relationship: residue selection

Determination of relevant residues in terms of the molecular activity of methyltranferases is a central objective given its importance to explain the converging and functional mechanisms of this extensive class of enzymes, particularly for those acting on RNA. Initially, the *E. coli* RsmG sequence (Uniprot id P0A6U5) was used to search the homologues distributed in the main bacteria groups for the purpose of studying the residue conservation in this family of proteins. After a Blastp search, we selected approximately 30 different proteins representing the major groups of bacteria, which were aligned by the MUSCLE application (see Methods). A multiple alignment of the RsmG proteins is shown in Figure 2. In order to determine the residues responsible for: i) AdoMet-binding, additionally to those conforming the cannonical G-X-G-X-G motif (Schubert et al., 2003); and ii) catalytic transference of the CH3 group, we performed a manual docking of the AdoMet cofactor in an apo-form of the *Escherichia coli* RsmG (PDB 1JSX). AdoMet-containing structural templates were searched by the Blastp algorithm in the PDB database using the HMM-RsmG sequence profile (see Methods) as bait. Consequently, the N^5-Glutamine Methyltransferase HemK from *Escherichia coli* (PDB id 1T43) proved to be the best match (≈24% identity). However, the 1T43 molecule was achieved with S-adenosyl-L-homocysteine (AdoHcy) instead of AdoMet. Notwithstanding, the 1NV8

structure corresponding to HemK from *Thermotoga maritima* and crystallised with AdoMet was also used to build a model of the *E. coli* RsmG coupled with its cofactor. Manual docking was done by superposing the canonical AdoMet binding site G-x-G-x-G (Figure 3) without allowing clashes.

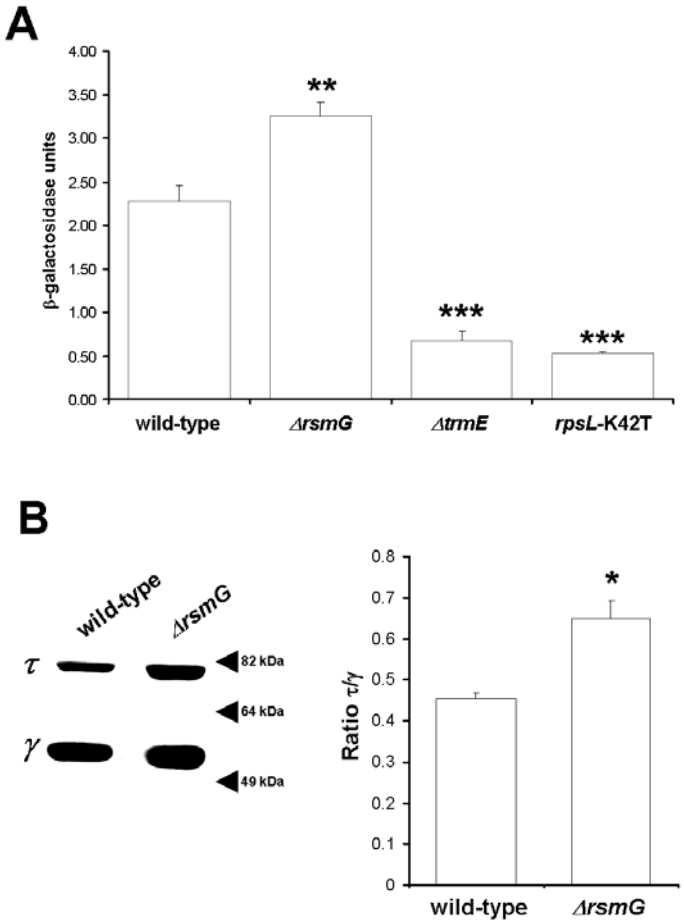

Figure 1. Read-through and frameshifting phenotypes of the *rsmG* mutant. A – β-galactosidase activity for the wild-type Dev16 and its *rsmG* mutant. The *ΔtrmE* and *rpsL*K42T mutations were studied as a control of this assay. ** p<0.010. *** p<0.001. B – Analysis of the isoform expression for the programmed *dnaX* frameshifting. Band densitometry and the τ/γ ratio were measured with four different replicates. * p<0.0175.

After fitting the AdoMet structure into the *E. coli* RsmG structure, we focused on residues with contact surfaces and different moieties of the cofactor to test their involvement in its binding by the experimental

approaches explained below. Simultaneously, we localised several well conserved residues in the *E.coli* RsmG structure, as depicted in the previously drawn multiple alignment. Only the G73, G75, and D96 residues had major contact surfaces with AdoMet (see Figure 3C). In addition, other residues were also observed in AdoMet binding, which are also likely to be involved. The L95, V72, and V124 residues in the *E. coli* RsmG make up a small hydrophobic pocket where adenine moiety of AdoMet is buried. Although these positions shown no conserved identity through multiple alignment, their hydrophobicity is strongly constrained (see Figure 1 - dots at top). Therefore, they could all be considered part of the AdoMet-binding site given their chemical conservation and structural arrangement.

Besides the residues involved in AdoMet binding, other residues which presumably accomplish both RNA binding and catalytic activities were also predicted. The R139 residue is a highly conserved position in all the RsmG sequences studied. This residue structurally appears close to the CH_3 donor group of AdoMet (see Figure 3C). In previous analyses of tRNA methyltransferases, a well conserved arginine residue was depicted to be directly involved in the transference of the methyl group by coordinating the guanine base ring to be modified (Elkins et al., 2003). Other reactive groups lying close to R139 were observed in the entry of a deep pocket with the P79 residue at the bottom. Consequently, H53 and D56 residues structurally converge with R139. Notwithstanding, N39 could also be considered according to sequence conservation, but its structural placement is uncertain as it does not appear in the original *E.coli* RsmG structure (PDB 1JSX). According to secondary structure data, N39 could be located in an unstrutured loop connecting the $\alpha 2$ and $\alpha 3$ helices. At this point, we predicted several residues which, taken together, would become catalytic requirements for the methyltranference carried out by RsmG. However, their importance for protein function remains to be individually determined.

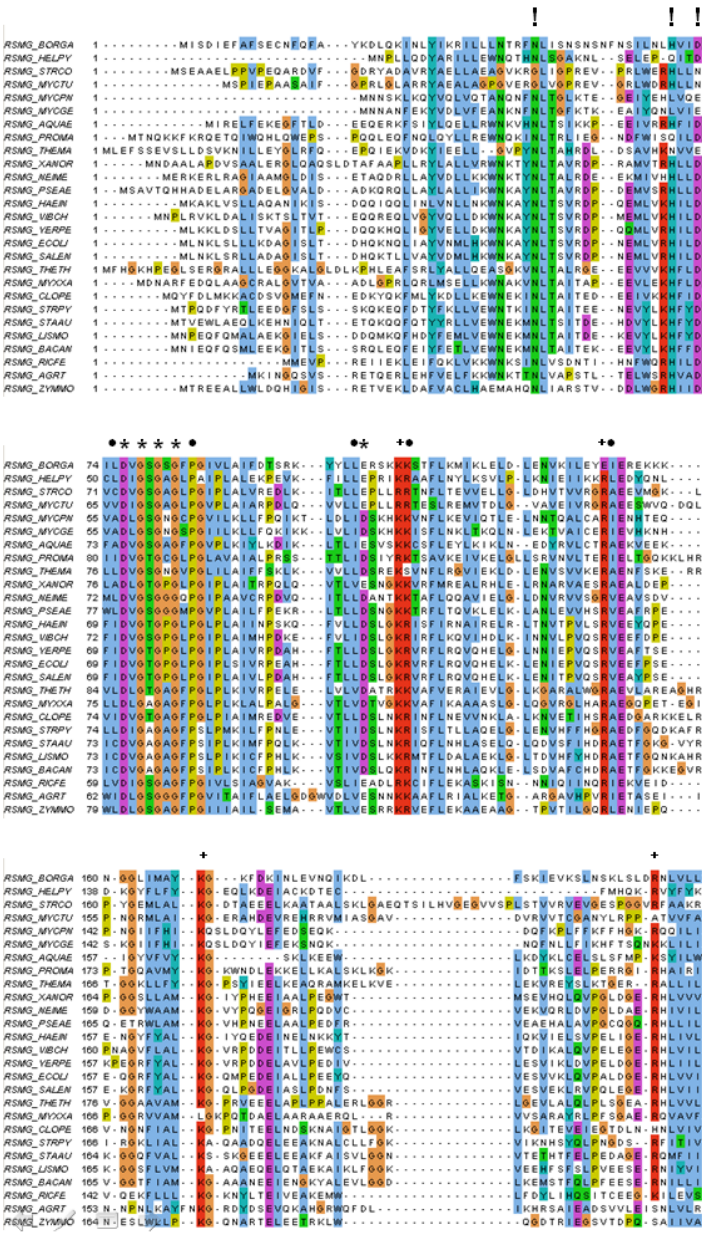

Figure 2. Sequence alignment of the RsmG family of proteins. (*) depicts the canonical motif for AdoMet binding; (●) denotes the additional residues interacting with AdoMet; (+) represents those residues probably involved in RNA binding; (!) denotes the residues involved in catalysis.

The third group of residues to be tested comprises those probably involved in RNA binding. These residues were predicted by directly inspecting the positively charged amino acid conservation. Then residues K100, R101, R123,

K165 and R197 (see Figure 2, "+" signs) seem to be well conserved in the whole alignment of the RsmG family and are localised on the same protein surface close to the other relevant residues detailed in our predictions (Figure 3C).

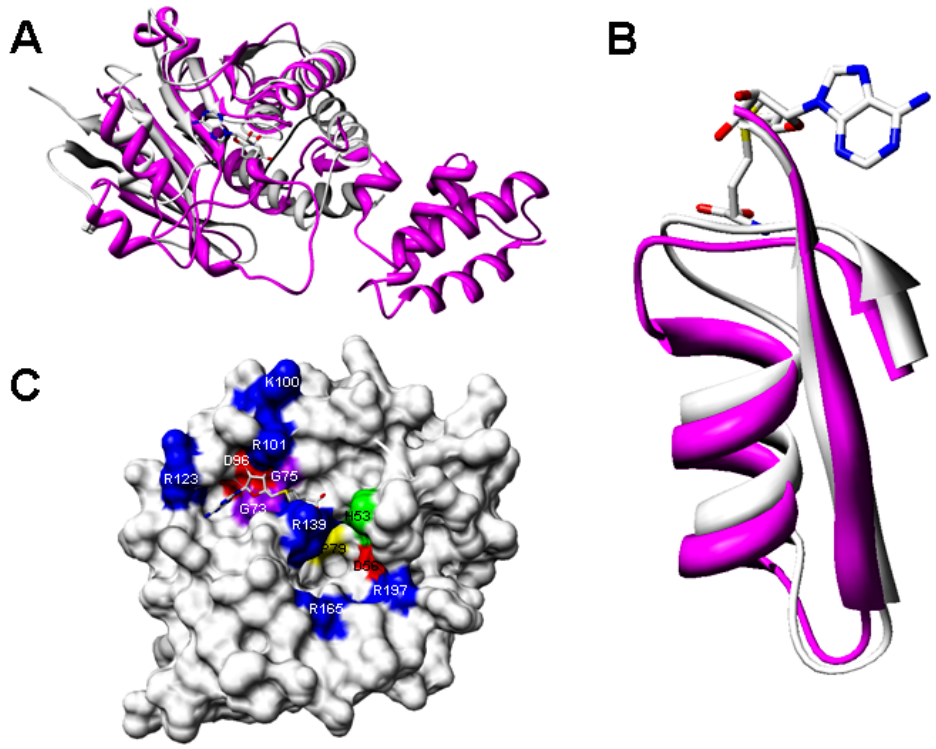

Figure 3. Structural analysis of the *E.coli* RsmG protein with manually docked AdoMet. A – Structural alignment of the 1JSX (*E. coli* RsmG) and 1NV8 (*T. thermophilus* HemK) proteins. B – Superposition of the canonical AdoMet binding motifs of both structures. C – Structural localisation of those residues selected to be tested for their functions.

LOW-LEVEL STREPTOMYCIN RESISTANCE

Given that *rsmG* mutants show low-level streptomycin resistance (Okamoto et al., 2007), the functionality of each RsmG mutant selected was tested with a simple assay to check growth ability in LB plates supplemented with 20 ug/mL of streptomycin. This *in vivo* analysis to test the restoration of streptomycin sentivity was done by using an expression system to allow a similar recombinant RsmG expression to that observed for native RsmG in a

wild-type strain. To obtain this condition, recombinant RsmG-Flag wild-type and mutants were cloned under the control of the P_{BAD} promoter (see Methods) and were inserted into *ΔrsmG* cells. RsmG native levels for all the recombinant proteins in *rsmG*-null cells were achieved without using an expression inducer, normally L-arabinose. The results of this analysis are shown in Figure 3. The expression level per mutant is expressed in terms of the native RsmG expression in the wild-type cells. Detection of this level of expression was only possible using polyclonal anti-RsmG (see Methods). The results of the remaining mutants were similar to those presented for the mutants listed in Figure 4 (data not shown).

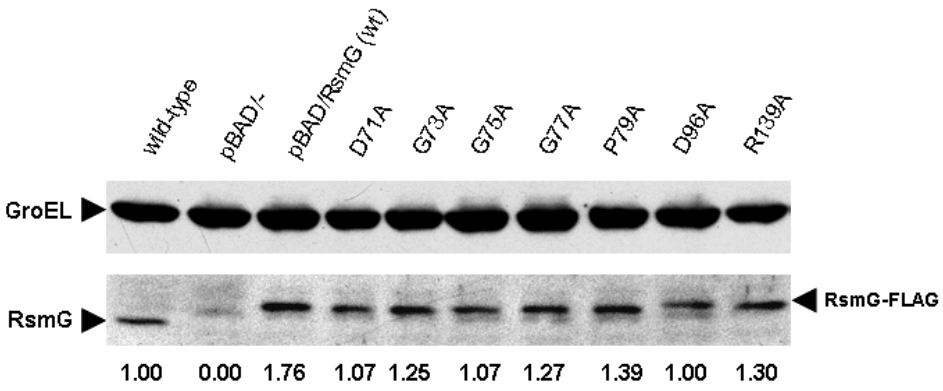

Figure 4. Expression analysis of the RsmG mutants to be tested.The numbers below the lanes correspond to the native RsmG fold expression. GroEL expression was used as a loading control.

The same culture conditions were used to test the low-level streptomycin resistance in each mutant. Th results of this assay are presented in Figure 5. Therefore, we observed that several mutants appear to be affected by rRNA modification.

In this way, mutants may be classiffied according to the level of growth as follows: i) null mutants with full growth as seen in *ΔrsmG* (empty plasmid): D71A, G75A, G77A, P79A, D96A, R139A, N39A, D56A, K100A-R101A and R197A; ii) mutants partially affected with attenuated growth: H53A and R139K; and iii) mutants sensitive to streptomycin resembling the wild-type phenotype: D73A, R123A and K165A. All the strains involving mutations in the canonical AdoMet binding motif, except G73A, showed a highly affected

function of the RsmG protein. Interestingly, and despite the G73 residue being seen to directly interact with AdoMet, it does not seem to be affected by Alanine replacing, possibly because hydrophobicity was relatively maintained. Notwithstanding, the same replacement in residues G75 or G77 completely abolished RsmG functionality. Since residues G75 and G73 offer the major contact surfaces for AdoMet binding, the most important residue where this interaction lies is apparently G75.

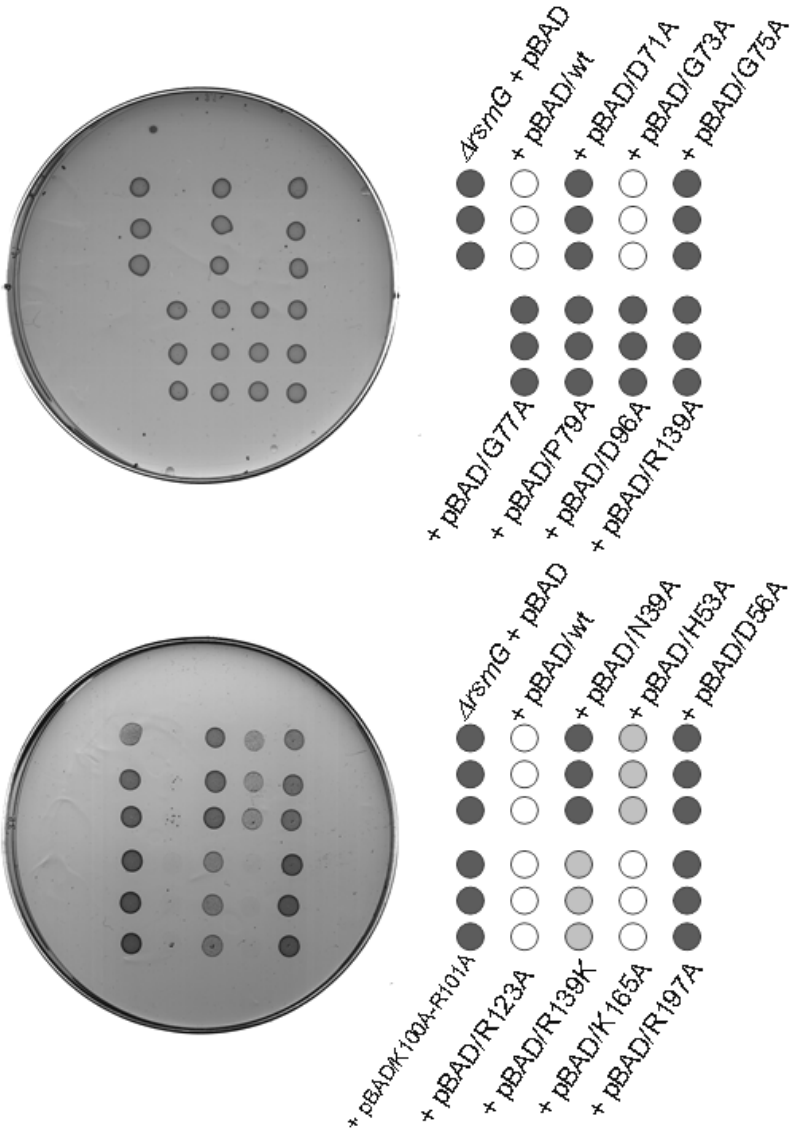

Figure 5. Low-level streptomycin resistance assay in RsmG mutants.

On the other hand, the D96A mutant was profoundly affected, as were other mutants of the AdoMet binding motif and, interestingly, the conserved P79 showed a null function phenotype. The presence of this residue within AdoMet binding motif may be associated with this function. However, P79 has no major interaction surface with AdoMet, but interacts with another important residue of the motif instead: D71.

Regarding the catalityc function, the residues predicted in this molecular task were affected to different extents. As a result, the main candidate, R139, showed complete loss-of-function when replaced in the Alanine scanning. Nonetheless, mutation R139K partially restored the function of this residue. This result supports the functional relevance of this residue in catalysis, where its positive charge would be necessary to coordinate the ribonucleotide target. Mutations N39A and D56A produced inactive versions of RsmG, for instance R139A, whereas H53A was partially active as R139K. Finally, those residues predited to be involved in RNA binding were functionally unaffected in the case of mutants R123A and K165A. However, double mutant K100A-R101A was completely affected, probably because of the double mutation itself. Surprisingly, the R197A mutation turned RsmG into an inactive version of the protein.

IN VIVO COMPLEMENTATION OF 16S RRNA METHYLATION

In order to do an in-depth study into the functional activity of each designed mutant, we analysed the modification status of *Escherichia coli* 16S rRNA from the *ΔrsmG* cells complemented with the different mutants obtained in this study. By reproducing same culture conditions of the low-level streptomycin assay, 16S rRNA was isolated from all the mutants and was analysed by HPLC after hydrolisation (Figure 6).

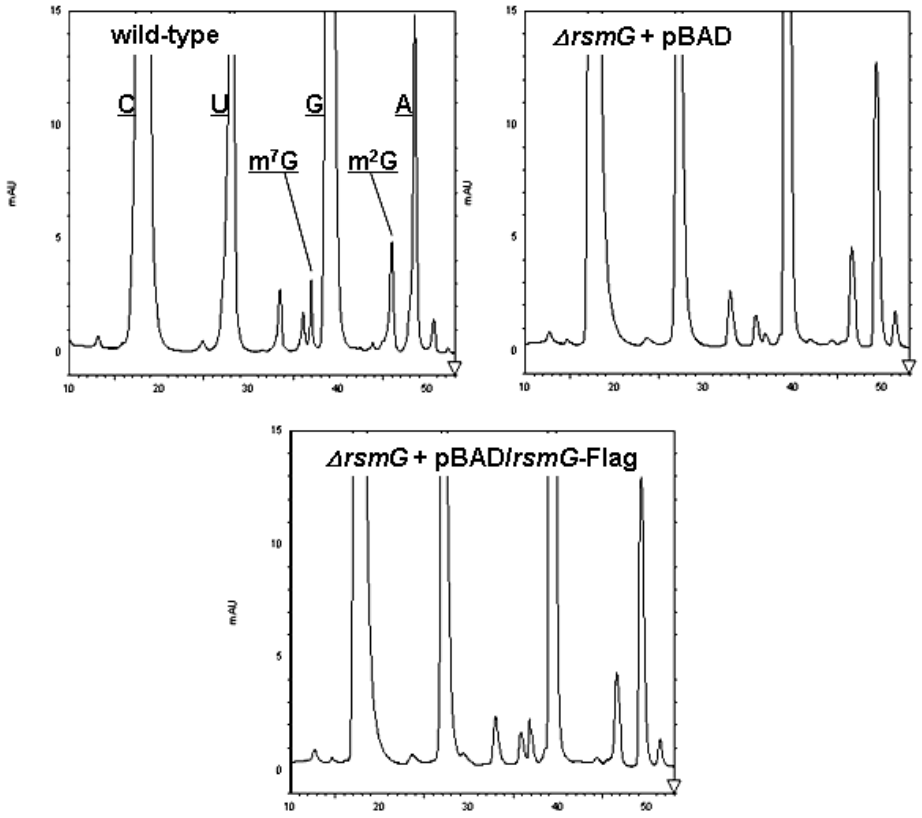

Figure 6. The HPLC profile of *Escherichia coli* 16S rRNA. A lack of m^7G nucleoside was evidenced in the *ΔrsmG* strain which was restored when complemented with recombinant *rsmG*-Flag cloned in the pBAD vector.

A lack of the m^7G nucleoside was evident in the *ΔrsmG* cells when compared with both the wild-type and the null mutant complemented with pBAD/*rsmG*-Flag. Then, the modification level of 16S rRNA was also studied in each mutant to detect a more specific activity per protein. The result of *in vivo* activity in all the mutants is offered in Table 2. With this functional assay, it is possible to obtain a variety of greater activities than in the low-level streptomycin resistance assay.

Table 2. Properties of the RsmG mutant proteins.

Protein	Predicted Function	Low-level Sm resistance	*in vivo* activity[a]	AdoMet affinity[b] μM
wild-type		-	1.00	0.39
N39A	Catalysis	++	0.24	1.19
H53A	Catalysis	+	0.28	0.77
D56A	Catalysis	++	0.13	0.47
D71A	AdoMet binding	++	0.00	0.04
G73A	AdoMet binding	-	0.80	1.39
G75A	AdoMet binding	++	0.01	9.65
G77A	AdoMet binding	++	0.00	4.09
P79A	AdoMet binding	++	0.05	4.38
D96A	AdoMet binding	++	0.00	ND
K100A/R101A	RNA binding	++	0.00	0.33
R123A	RNA binding	-	0.84	ND
R139A	Catalysis	++	0.00	0.32
R139K	Catalysis	+	0.57	ND
K165A	RNA binding	-	0.90	ND
R197A	RNA binding	++	0.15	ND

a) *In vivo* activity per protein is reported and normalised to the wild-type. The data are the mean of two replicates of the m^7G/m^2G peak area comparison obtained by HPLC analysis (see Methods). b) AdoMet affinit per protein is shown as the mean of two independent experiments where respective *kd* constants were determined. ND = not determined.

Most of the *in vivo* activities of the mutants involved in AdoMet binding resulted in a complete loss-of-function given that no trace of m^7G modification was observed in their 16S rRNA. Notwithstanding, the apparently unaffected G73A mutation retained 80% of wild-type activity. The activity of those mutants involving the presumably catality residues show how their functions are greatly affected by less than 30% of wild-type activity. The strongest effect was observed in mutant R139A where no trace of m^7G modification was evidenced, whereas partial restoration of activity was observed in R139K, which is in agreement with the previous streptomycin resistance data.

Mutants R197A and K100A/R101A show no *in vivo* activity in the set of candidates. Interestingly, the R197 residue seems most RNA binding critical for the protein function.

AdoMet affinity

The study of AdoMet affinity offers functional insights for the residues studied here. By taking into account that wild-type RsmG shows an AdoMet affinity mean of 0.39 μM, the major changes in this parameter were observed in those residues involved in its binding. Thus, all the mutants of the canonical AdoMet binding motif showed loss-of-affinity for this cofactor (Figure 7). The largest effect was observed in G75A where AdoMet *kd* increased by more than 24-fold compared to that seen in wild-type RsmG (Table 2). Similarly, mutants G73A and G77A were also seen to decrease in AdoMet affinity. Notwithstanding, the G73A affinity was only 3.5-fold less than that observed in wild-type RsmG, which generates a protein with approximately 80% of the activity (Table 2). On the other hand, the D71A mutation conferred more affinity to RsmG by increasing its *kd* for AdoMet up to 10-fold. In this way, the null activity of this mutant may be possibly explained by its incapacity to hydrolise the AdoMet.

This scan for AdoMet binding property enabled us to detect a new residue involved in AdoMet binding. Consequently, when the well conserved P79 residue was mutated to alanine, the RsmG protein decreased its affinity for AdoMet by more than 11-fold. Then RsmG D96A showed structural unstability since it was not possible to obtain a soluble HIS-tagged recombinant protein by superproduction. Thus, we cannot rule out that the null *in vivo* activity observed by HPLC analysis is the result of misfolding caused for alanine mutation.

The affinity value obtained from the R139A mutant demonstrates that this residue did not affect cofactor binding, although its functionality was severely impaired. The AdoMet affinity of the remaining residues predicted to be involved in the catalysis varied slightly compared to wild-type RsmG. Consequently, it is likely that they accomplish the molecular function referred to here. For the double mutant K100A/R101A, AdoMet binding was not affected and its null *in vivo* activity could be the result of both residues playing a relevant role in RNA binding.

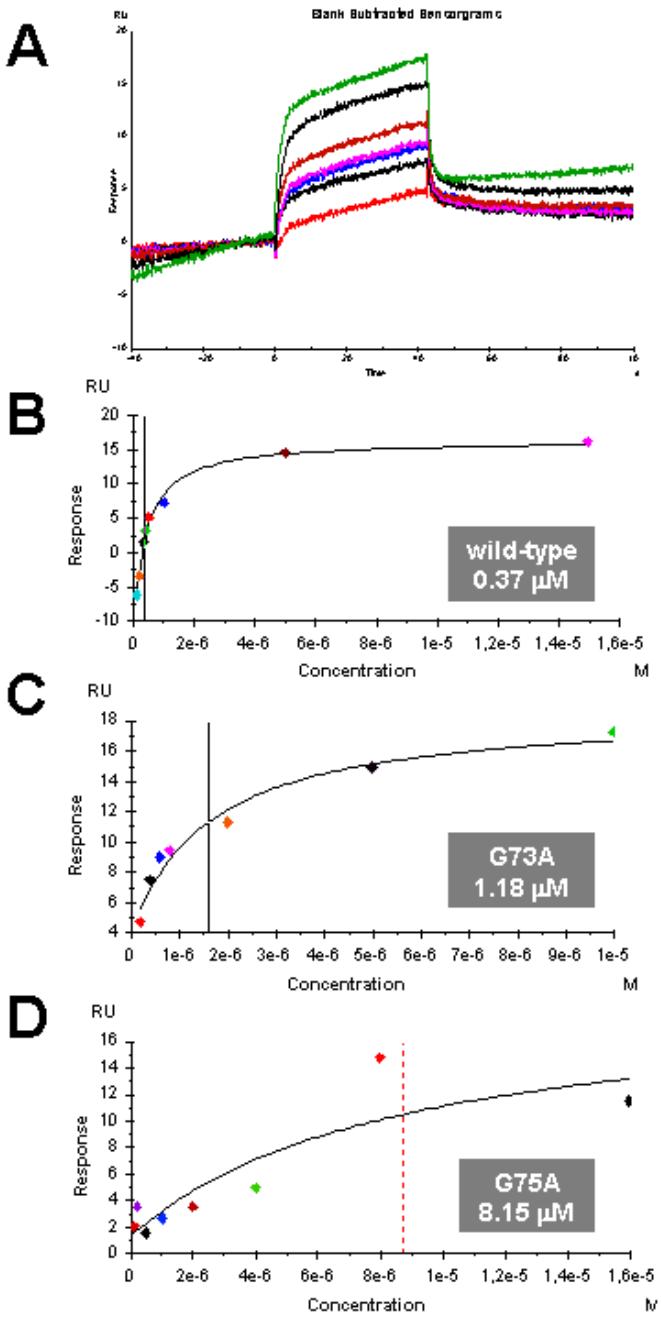

Figure 7. AdoMet affinity of RsmG proteins by Surface Plasmon resonance. A – Sensogram of AdoMet binding for wild-type RsmG. B, C and D – Adomet saturation curve for the wild-type, G73A, and G75 RsmG recombinant proteins.

DISCUSSION

Important biochemical and functional insights into RsmG methyltransferase and the chemical modification that it performs in position G527 of rRNA 16S in *Escherichia coli* are provided here. We have found that when this modification is lacking, there are different effects on translation. Firstly, we report an error-prone phenotype in the *ΔrsmG* by increasing the read-through rate of the UAG stop codon in the reporter *lacZ*. Besides, involvement in the reading frame control was also observed because absence of m^7G527 slightly changes the expression pattern in the *dnaX* τ and γ isoforms. Such changes in expression patterns of the *dnaX* τ and γ isoforms has been also studied in the context of the modifications occurred in the wooble position of tRNAs (Tsuchihashi & Brown, 1992). Despite the *ΔrsmG* mutation being previous characterised with no different read-through rate than that observed in wild-type cells (Okamoto et al., 2007), we could rescue a visible effect of this mutation by using an activator of the *lacZ* expression.

In the rRNA methyltransferases context, this is a pioneering report of the translational effects of the *ΔrsmG* mutation. Even with the most extensively studied rRNA methyltransferase, RsmA, no certain effects in ribosome fidelity have been disclosed to date. We describe how a highly conserved modification carried out by *rsmG* is directly involved in ribosome fidelity, and controls the appropriate read-through rate and the ribosome reading frame.

In addition to the effects in translation, the biochemical and functional characterisation of the RsmG protein has enabled us to describe a broad set of residues with relevant molecular roles such as AdoMet binding, RNA interaction, and even catalysis. Consequently, we describe the relevance of different residues confined to the canonical AdoMet binding site where G75 is presented as the core residue for the cofactor attachment, which is argued by AdoMet affinity undergoing a >24-fold decrease when replaced with alanine (Table 2). At the same time, we found that conserved P79 is also involved in AdoMet binding since mutation P79A decreases AdoMet affinity by more than 10-fold. Involvement of RNA binding was hypothesised for both the conserved and positively charged residues K100, R101, R123, K165 and R197. Notwithstanding, only R197A and mutations K100A/R101A severely impaired the *in vivo* activity of RsmG, as demonstrated in low-level streptomycin resistance and in the *in vivo* rRNA modification assays. Then

their role in the interaction with RNA is still probable since AdoMet affinity is not affected, at least in the K100A/R101A mutant (Table 2). Interestingly, the R197 residue comes closer to the active site following the structural arrangement of the protein after AdoMet binding, according to very recent strcutural analyses (Gregory et al., 2009). Therefore, a more critical role could be expected for R197.

Those residues predicted to be possibly involved in catalysis are located near the AdoMet binding site. However, their involvement in cofactor binding has been discarded since AdoMet affinity is not severily affected (Table 2). Simultaneously, the *in vivo* activities of rRNA modification are poor. In terms of the data recopiled for them, we propose that residues R139, H53, and D56 must be implicated in the catalysis by coordinating the target nucleoside. Similar residues have been postulated to allow the m^1G modification in tRNAs by TrmD (Elkins et al., 2003). In this way, we postulate that the mayor role in the G527 coupling would be played by R139 because of its high evolutive conservation in the RsmG family sequence (Figure 2).

Furthermore, the N39 residue has been disclosed to be important for *in vivo* activity. The localisation of this residue is not fully understood because other neighbour residues, which probably make up an unstrutured loop which is difficult to cristalise, are missing in all the RsmG structures. Thus it seems that this loop could be responsible for the conformational change in the protein after AdoMet binding. This hypothesis is supported by N39A showing a slight loss of cofactor affinity, which is 3-fold higher than that seen in wild-type RsmG.

Globally, we present a broad set of central residues for the RsmG methyltransferase function. Essentially, these residues are responsible for AdoMet binding and catalysis. Although some form part of the structural core which is common to methylases, their chemical nature and probable assigned roles will prove useful to study CH_3 transference mechanisms in other RNA methyltransferases given their evolutive convergence.

Acknowledgements

The author wishes to thank to National BioResource Project (NBRP) Japan and its KEIO collection for donating *ΔrsmG*; Magda Villaroya for her technical assistance in HPLC procedures; and every member of the Molecular Genetics Laboratory at CIPF for their constructive suggestions while accomplishing this research work.

Chapter 3
YibK/TrmL is the 2'-O-methyltransferase that modifies the wobble position in *Escherichia coli* $tRNA^{Leu}$ isoacceptors

Summary

The introduction of specific post-transcriptional modifications in tRNAs is a necessary step for the correct functioning of the translation process. Despite much research, the set of enzymes responsible for the modifications found in bacterial tRNAs has not been completely identified. To help bridge this gap, we applied a comparative genomics approach to search for the novel *Escherichia coli* genes involved in tRNA modification. Our results show that tRNA modifying enzymes share a specific genomic and evolutionary context with not only those genes known to be involved in rRNA modification and ribosomal architecture, but with a set of 15 genes of an unknown function. Based on this observation and on

additional sequence analyses, we predict that a subset of these uncharacterised genes is likely to be involved in tRNA modification. The implication in this process of 11 of these genes was further studied by experimental screening. Several mutants of our candidate genes show a hyperaccuracy phenotype which is indicative of a role in translation. Consistently with our predictions, and by using an RNA mass spectrometry approach, we were able to identify how *yibK* (now known as *trmL*) is responsible for ribose methylations in the wobble position in $tRNA^{Leu}_{CmAA}$ and $tRNA^{Leu}_{cmnm5UmAA}$. Deletion of *yibK* does not affect exponential growth in rich media, but confers a selective disadvantage in competition with isogenic wild-type cells, meaning that the loss of the YibK function involves a biological cost.

INTRODUCTION

Most non-coding RNAs (ncRNAs) are post-transcriptionally modified. Nucleoside modifications usually reinforce a particular tertiary structure and are, therefore, required for the proper functioning of RNA. In particular, post-transcriptional nucleoside modifications of transfer-RNAs (tRNAs) have been shown to stabilise their characteristic L-shaped structure, to modify the cognate codon recognition, to affect aminoacylation, and to stabilise codon-anticodon wobble base pairing, thus preventing ribosomal frameshifting (Björk & Hagervall, 2005b). Such post-transcriptional modifications are carried out by a number of specific enzymes, whose activity has been shown to be critical for the proper decoding of genetic information. In some cases, the lack of tRNA modifications leads to a misreading of mRNA (Bregeon et al., 2001; Nasvall et al., 2004). Given their involvement in the global translation process, tRNA modification pathways are important to all cellular processes including pathogenicity and different cell stress responses (Zimmermann et al., 1973; van Buul et al., 1984; Karita et al., 1997; Forsyth et al., 2002; Sha et al., 2004).

Recent genomic and post-genomic (proteomics and transcriptomics) analyses have identified many new genes that encode enzymes which are responsible for RNA modifications (Bishop et al., 2002; Reader et al., 2004; Grosjean et al., 2008; El Yacoubi et al., 2009). Yet despite such efforts, the enzymes responsible for certain RNA modifications remain unidentified. For instance, the enzyme synthesising the m^2G1516 modification present in the

16S rRNA of *Escherichia coli* is unknown. Regarding 23S rRNA, 22 of the 25 modified nucleotides have been reported to be ascribed to a particular modifying enzyme (briefly reviewed in (Purta et al., 2009)). Among the non-coding RNAs, tRNAs have the highest density of modifications per molecule (approximately 10% of nucleotides present in a tRNA molecule may be modified post-transcriptionally). Cells invest substantial biosynthetic effort and resources to the post-transcriptional modification of RNA. As a result, approximately 1% of bacterial genome is invested in the tRNA modification process (Björk & Hagervall, 2005b). Notwithstanding, such a calculation may be underestimated since many of the enzymes involved in the biosynthesis of donor groups can also be considered to belong to the tRNA-modification pathway (i.e., the *aroABC* operon, genes for THF biosynthesis, genes for the Fe-S cluster assembly). Additionally, the synthesis of hypermodified nucleosides in tRNA, with the presence of bulky groups attached to purine or pyrimidine rings, requires a multi-enzyme pathway (Czerwoniec et al., 2009). Accordingly, certain pathways have been proposed to proceed in several steps, involving serial modifications by different enzymes (Hagervall et al., 1987; Ikeuchi et al., 2006; Lundgren & Bjork, 2006; Moukadiri et al., 2009). Therefore in bacteria, such as *E. coli* and *Salmonella spp*, there are more protein-encoding genes expected to be involved in tRNA modification than the modifications known. The presence of 46 different modified nucleotides in bacterial tRNA has been reported (the RNA Modification Database, http://library.med.utah.edu/RNAmods/). Thirty-one of them are found in tRNAs from *Escherichia coli*, whereas 27 genes have been described to participate in some way in tRNA modification (Björk & Hagervall, 2005b). Although the number of genes involved in tRNA-modification has increased in recent years (Bujnicki et al., 2004; Purta et al., 2006; Ikeuchi et al., 2008; El Yacoubi et al., 2009; Golovina et al., 2009), several are still missing.

In order to identify the additional genes coding for the enzymes involved in tRNA modification, we have adopted herein a comparative-genomics approach that exploits the availability of a high number of fully-sequenced bacterial genomes. Comparative genomics offers new approaches to predict the function of genes through bioinformatics analyses (Gabaldon & Huynen, 2004; Gabaldon, 2008). These include, among many others, the use of phylogenetic profiles (Pellegrini et al., 1999) that infers functional interactions among proteins based on their correlated evolution reflected by the presence or absence profile obtained from a set of known genomes. Other

types of approaches are based on gene clustering (Overbeek et al., 1999; Zheng et al., 2002), gene fusion (Snel et al., 2000a; Yanai et al., 2001), co-expression (Stuart et al., 2003), etc. In the present study we apply a combination of several of the above-mentioned techniques to find the candidate genes that share similar "genomic contexts" with known tRNA modification enzymes. We report a list of 15 genes which are likely to be involved in tRNA modification in accordance with to their shared genomics contexts and domain composition. The implication in this process of 11 of these genes has been further studied by experimental analyses. Our results demonstrate that one of our candidate genes, *yibK*, is responsible for the ribose methylations which are characteristic of wobble positions in the $tRNA^{Leu}_{CmAA}$ and $tRNA^{Leu}_{cmnm5UmAA}$. Simultaneously, we detected a co-evolution among the different components of translation machinery because of the interactions established there. Consequently, we expect some of our candidates to be involved in other translation tasks other than tRNA modification.

METHODS

COMPARATIVE GENOMICS - BIOINFORMATICS PREDICTIONS

SEQUENCE DATA

A list of known tRNA modification enzymes (Table 1) and their respective protein sequences was obtained from Björk, 2005 (Björk & Hagervall, 2005b), along with the proteins recently identified as tRNA-modifying (Bujnicki et al., 2004; Purta et al., 2006; Ikeuchi et al., 2008; El Yacoubi et al., 2009; Golovina et al., 2009). The proteins encoded in completely-sequenced bacterial genomes were downloaded from the Integr8 database at EBI (Kersey et al., 2005) as of January 2007.

GENERATION OF PHYLOGENETIC PROFILES

Smith-Waterman searches were run using sequences from known tRNA modification enzymes as a query against the above-mentioned database of complete bacterial proteomes. One particular gene was considered to be present in a given species whenever it produced a hit with an e-value lower than 10^{-3}, which aligned over 50% of the query sequence. Phylogenetic profiles were represented as matrices of 0's and 1's, indicating presences or

absences, respectively. Distances between profiles were computed using the Hamming Distance.

Analysis of gene fusion and chromosomal neighbourhood

The gene neighbourhood analysis and the search for gene fusion events in other genomes were carried out in the STRING webserver (von Mering et al., 2007). For the purpose of obtaining more reliable predictions of protein interactions, we fitted the confidence scoring at ≥ 0.600.

Bacterial strains

All the knockout mutants of the candidate genes identified by comparative genomics, as well as the *mnmE* mutant, were obtained from the Keio collection (Baba et al., 2006). The *mnmG* mutant carrying a Tn10 insertion was kindly donated by Dr. Bregeon (Bregeon et al., 2001). Using phage P1 transduction (Miller, 1990), null-mutations of candidate genes were introduced into strain IC4639 (Yim et al., 2006), a wild-type derivative from strain Dev16 (Elseviers et al., 1984). The *miaA* mutant was kindly donated by Dr. Glen Bjork. The correct insertion of mutations was checked by PCR using the primers flanking the replaced gene and the internal primers for the kan^r gene (Datsenko & Wanner, 2000), or other insertion elements used to knockout genes of interest.

Read-through assay

β-galactosidase activity was determined according to the Miller procedure (Miller, 1990) for all the null mutations transduced into the wild-type Dev16 strain (see above). A simple variation of the assay was done in the overnight and day cultures, which were supplemented with 1mM IPTG (Apollo Scientific Ltd) and 0.3 mM cAMP (Sigma) in order to increase β-galactosidase activity. The activity values for all the strains were examined for statistical significance which was achieved from at least three independent replicates analysed by a t-test and Welch's correction.

Growth and competition experiments

The growth rate was calculated as doubling the time of each strain culture at the steady-stated log phase by linear regression. Competition experiments were carried out as previously reported (Gutgsell et al., 2000). Briefly, wild-types and mutants were grown separately in the stationary phase by incubation at 37ºC. Equal volumes of wild-type and individual mutant

cultures were mixed and a sample was immediately taken to count the viable cells on LB plates with and without the antibiotic required to estimate the mutant cell content. Six 24-hour growth cycles at 37ºC were performed by diluting the mixed cultures to 1/1000 on LB media; each cycle corresponded to ten or eleven cell divisions. After the sixth cycle, the mixed culture was analysed for its wild-type:mutant cell content, as before.

Nucleoside HPLC-profiles

Bacterial strains were grown overnight in medium LB. Then, cultures were diluted 100 times in 100 mL of LB media and grown to OD_{600} 0.7 to 0.8 units. Cells were harvested by centrifugation, and pellets were resuspended in 0.4 mL Buffer-A (25 mM Tris pH 7.4, 60 mM KCl, 10 mM $MgCl_2$). Next 2 mg of Lysozyme (Sigma) were added and the suspension was incubated at 37ºC for 15-20 minutes. Cell suspension was lysed by three-time freeze-thaw using liquid nitrogen. Then 0.6 mL of Buffer-B (Buffer-A supplied with 0.6% Brij35, 0.2% Na-deoxycholate, 0.02% SDS) and 0.1 mL of Phenol (citrate equilibrated, pH 4.3) were added and mixed. The suspension was incubated on ice for 15 minutes and the aqueous phase was twice extracted with phenol 1:1. RNA was precipitated by 2.5 volumes of cold 99.5% ethanol containing 1% (w/v) potassium acetate. After washing the pellet once with 70% ethanol, it was dissolved in 2 mL buffer R200 (100 mM Tris-H_3PO_4 at pH 6.3, 15% ethanol, 200 mM KCl) and placed in a Nucleobond AX500 column (Macherey-Nagel) to be pre-equilibrated with 10 mL of the same buffer. The column was washed once with 6 mL R200 and once with 2 mL R650 (R200 supplied with 650 mM KCl instead of 200 mM). tRNA was eluted with 6 mL R650. tRNA was precipitated by 0.7 volumes isopropanol, washed once with 70% ethanol, and dissolved in water. Afterwards, 50ug of the tRNA mixture were hydrolysed by nuclease P1 (Sigma) overnight incubation in water with 1mM $ZnSO_4$, followed by treatment with *Escherichia coli* alkaline phosphatise (Sigma) at pH 8.3 for 2 hours. The hydrolysate was analysed by HPLC using a Develosil 5µ RP-AQUEOUS C-30 column (Phenomenex®) with gradient elution to obtain optimal separations of nucleosides. Buffer A had 2.5% methanol, 10mM $NH_4H_2PO_4$, pH 5.1, while buffer B had 25% methanol, 10mM $NH_4H_2PO_4$, pH 5.3. The time for gradient elution was extended to 100 minutes. All the HPLC-nucleoside mutant profiles were compared with those derived from the wild-type. Approximately, 16 predominant and well-known (by UV spectra according to (Gehrke & Kuo,

1989)) tRNA modifications seen in the wild-type strain were evaluated to be absent in the mutants of the candidate genes at 254 nm.

RNA Mass Spectrometry

Bulk tRNA from wild-type *ΔyfiF*, and *ΔyibK* was isolated as above, and 2500 pmol were incubated overnight with 85 mM 3-HPA, 1 unit of RNase T1 (USB) or 3 ug of RNase A at 37ºC. Digested tRNA was mixed at 2:1 with 1 M triethylammonium acetate (TEAA) and loaded onto a microcolumn for reverse-phase type chromatography on a GELoadertip containing the Poros R3 matrix (Applied Biosystems) to be pre-equilibrated with 10 mM TEAA. After washing twice with 10 mM TEAA and once with 10% acetonitrile and 10 mM TEAA solution, the larger fragments were eluted with a solution of 30% acetonitrile and 10 mM TEAA. Samples were dried and dissolved in 3 ul of H_2O prior to analysis in a MicroMass MALDI-Q-TOF Ultima Mass Spectrometer, recording in the positive ion mode (Kirpekar et al., 2000; Douthwaite & Kirpekar, 2007).

In vivo complementation of modification in $tRNAs^{Leu}$

The *yibK* open reading frame from *Escherichia coli* was amplified using the following oligonucleotides: 5' - CGC**CCATGG**GT*CATCATCACCAT-CACCAT*ATGCTAAACATCGTACTTTACGAACCAGAAATTCCG and 3' – GCC**GGATCC**CTAATCTCT-CAATACCGCTCCCGG with the NcoI and BamHI restriction sites, respectively. The PCR product was digested and inserted into a NcoI/BamHI linearised pET15b plasmid by overnight incubation with T4 ligase at 16ºC. The pET15b/his-yibK construct was inserted into the BW25113 Δ*yibK* strain; empty pET15b plasmids were inserted into wild-type and Δ*yibK* cells as a control. Bulk tRNA isolation and mass spectrometry were done as described above.

Results

Selection of candidate genes

The genomic context of all the known *Escherichia coli* tRNA modification genes (Table 1) was analysed by comparing their phylogenetic profiles across 300 different genomes (see Material and Methods) and by using the STRING server (von Mering et al., 2007). Those proteins with a significant shared gene neighbourhood, similar co-occurrence or common gene fusion events,

were further considered (STRING score ≥ 0.6). Proteins sharing a genomic context with known tRNA modification enzymes included 15 proteins of an unknown or imprecise function (Table 2), as well as others known either to be components of the ribosome or to participate in other aspects of the translation process (Figure 1). This indicates a tight co-evolution between the tRNA modification pathways and the translation machinery, and suggests that our approach may have selected some candidates that participate in other aspects of the translational process and which are not merely involved directly in tRNA modification.

Table 1. tRNA-modifying enzymes used as searching bait.

Protein	Alternative gene name	Function description	Modification (position)	Reference	SwissProt Identifier
ThiI	*yajK*	Thiamine biosynthesis protein ThiI	S^4U (8)	(Mueller et al., 1998)	P77718
TruD	*ygbO*	tRNA pseudouridine synthase D	Ψ (13)	(Kaya & Ofengand, 2003)	Q57261
DusA	*yjbN*	tRNA-dihydrouridine synthase A	D (16, 17, 20, 20a)	(Bishop et al., 2002)	P32695
DusB	*yhdG*	tRNA-dihydrouridine synthase B	D (16, 17, 20, 20a)	(Bishop et al., 2002)	P0ABT5
DusC	*yohI*	tRNA-dihydrouridine synthase C	D (16, 17, 20, 20a)	(Bishop et al., 2002)	P33371
TrmH	*spoU*	tRNA guanosine-2'-O-methyltransferase	Gm (18)	(Persson et al., 1997)	P0AGJ2
TtcA	*ydaO*	tRNA 2-thiocytidine biosynthesis protein	S^2C (32)	(Jager et al., 2004)	P76055
IscA	*yfhF*	Iron-sulphur cluster assembly protein	ms^2i^6A (37)	(Leipuviene et al., 2004)	P0AAC8
MnmA	*ycfB, trmU*	tRNA-specific 2-thiouridylase	S^2U (34)	(Kambampati & Lauhon, 2003)	P25745
MnmE	*thdF, trmE*	tRNA modification GTPase	$cmnm^5U$ (34)	(Elseviers et al., 1984)	P25522
MnmG	*gidA*	tRNA uridine 5-carboxymethylaminomethyl modification enzyme	$cmnm^5U$ (34)	(Bregeon et al., 2001; Yim et al., 2006)	P0A6U3
SelU	*ybbB*	tRNA 2-selenouridine synthase	Se^2U (34)	(Wolfe et al., 2004)	P33667
SelD	*fdhB*	Selenide, water dikinase	Se^2U (34)	(Leinfelder et al., 1990)	P16456
Tgt	*vacC*	Queuine tRNA-ribosyltransferase	Q (34)	(Okada et al., 1979)	P0A847
QueA	*tsaA*	S-adenosylmethionine: tRNA ribosyltransferase-isomerase	Q (34)	(Slany et al., 1993)	P0A7F9
QueE	*ygcF*	7-cyano-7-deazaguanosine (PreQ0) biosynthesis protein	Q (34)	(Reader et al., 2004)	P64554
QueC	*yvaX*	Queuosine biosynthesis protein	Q (34)	(Gaur & Varshney, 2005)	P77756
QueF	*yqcD*	NADPH-dependent 7-cyano-7-deazaguanine reductase	Q (34)	(Van Lanen et al., 2005)	Q46920
TilS	*yaeN*	tRNA(Ile)-lysidine synthase	K^2C (34)	(Soma et al., 2003)	P52097
TadA	*yfhC*	tRNA-specific adenosine deaminase	I (34)	(Wolf et al., 2002)	P68398
MiaA	*trpX*	tRNA delta(2)-isopentenylpyrophosphate transferase	I^6A (37)	(Caillet & Droogmans, 1988)	P16384
MiaB	*yleA*	(Dimethylallyl)adenosine tRNA methylthiotransferase	ms^2i^6A (37)	(Esberg et al., 1999)	P0AEI1
TrmD		tRNA (guanine-N(1)-)-methyltransferase	m^1G (37)	(Bystrom & Bjork, 1982)	P0A873

TruA	*asuC, hisT*	tRNA pseudouridine synthase A	Ψ (38, 39, 40)	(Kammen et al., 1988)	P07649
TrmB	*yggH*	tRNA (guanine-N(7)-)-methyltransferase	m^7G (46)	(De Bie et al., 2003)	P0A8I5
TrmA	*rumT*	tRNA (uracil-5-)-methyltransferase	m^5U (54	(Ny & Bjork, 1980)	P23003
TruB	*yhbA*	tRNA pseudouridine synthase B	Ψ (55)	(Gutgsell et al., 2000)	P60340
TruC	*yqcB*	tRNA pseudouridine synthase C	Ψ (65)	(Del Campo et al., 2001)	P0AA41
CmoA	*yecO*	tRNA (cmo5U34)-methyltransferase	$mcmo^5U$ (34)	(Nasvall et al., 2004)	P76290
CmoB	*yecP*	tRNA (mo5U34)-methyltransferase	mo^5U (34)	(Nasvall et al., 2004)	P76291
MnmC	*yfcK*	tRNA 5-methylaminomethyl-2-thiouridine biosynthesis bifunctional protein	mnm^5U (34)	(Hagervall & Bjork, 1984a; Bujnicki et al., 2004)	P77182
TrmJ	*yfhQ*	tRNA (cytidine/uridine-2'-O-)-methyltransferase	Cm/Um (32)	(Purta et al., 2006)	P0AE01
TmcA	*ypfI*	tRNA N4-acetylcytidine synthase	ac^4C (34)	(Ikeuchi et al., 2008)	P76562
RimN	*yrdC*	tRNA threonylcarbamoyladenosine synthase	t^6A (37)	(El Yacoubi et al., 2009)	P45748
TrmN6	*yfiC*	tRNA (adenine-N(6)-)-methyltransferase	m^6A (37)	(Golovina et al., 2009)	P31825

In order to prioritise the list of candidate genes and to help develop testable hypotheses about the possible functions of the encoded proteins, we studied their domain composition using the SMART server (Schultz et al., 1998; Letunic et al., 2004). We also studied the related literature and database annotations (Figure 2). For those with homologues in eukaryotes, we investigated whether human homologues were identified in mitochondria by proteomic analyses according to MitoP2 (Elstner et al., 2009) or if they presented a putative mitochondrial localisation, as predicted by mitoprot (Claros, 1995). Following these criteria, we selected 3 specific genes (*hflX, ychF, and yhbZ*) which were highly conserved from bacteria to eukaryotes, whose products were apparently related with ribosomes and, in some cases (YhbZ), exhibited helper activity for the rRNA modifying enzymes (Teplyakov et al., 2003; Lamb et al., 2007; Shields et al., 2009; Wu et al., 2009). In addition to the above proteins, we highlight the presence of the highly conserved Era (RbaA) protein that contains an RNA-binding domain. This last feature is also observed in YbeZ (Figure S1).

Table 2. Candidate genes to be involved in tRNA modification.

Protein	Molecular Function Electronically Inferred (Database)	SwissProt Identifier
YqcC	None	Q46919
YcfC	None	P25746
YchF	GTP-dependent nucleic acid-binding protein engD (Swiss-Prot, Pfam)	P0ABU2
YajC	Preprotein translocase subunit (Pfam)	P0ADZ7

HflX	GTPase of an unknown function (Swiss-Prot, Pfam)	P25519
YbeY	Putative metalloprotease (Swiss-Prot)	P0A898
YbeZ	PhoH-like predicted ATPase that is induced by phosphate starvation (Swiss-Prot, Pfam)	P0A9K3
YnbZ	GTP binding protein belonging to the OBG family (Swiss-Prot, Pfam)	P42641
YggL	None	P38521
YhbC	None	P0A8A8
YraL	Possible methyltransferase (Pfam)	P67087
Era/RbaA	GTPase of an unknown function (Swiss-Prot)	P06616
EngC	GTPase that catalyses rapid hydrolysis of GTP with a slow catalytic turnover (Swiss-Prot)	P39286
YfiF	Uncharacterised tRNA/rRNA methyltransferase yfiF	P0AGJ5
YibK	Uncharacterised tRNA/rRNA methyltransferase yibK	P0AGJ7

Other specific highly associated domains with RNA modifications, such as the SpoU domain (in YfiF and YibK; see Figure 2) and the Pfam-DUF446 domain found in pseudouridine synthases (see protein YqcC in Figure S1), were also seen. Unlike *hflX*, *ychF, yhbZ*, and *rbaA,* there were other candidate genes showing less evolutionary conservation. In most cases, the presence of such genes was restricted to specific bacterial groups. YbeY is a metalloprotein which, in *E. coli*, has been involved in translation (Zhan et al., 2005; Rasouly et al., 2009), whereas YhbC (RimP) and EngC (RsgA) seem to participate in the 30S subunit maduration (EngC has also been proposed to be a translation regulator) (Kimura et al., 2008; Nord et al., 2009). YcfC is a membrane-associated protein of an unknown function (Green et al., 1996), whereas YajC is a subunit of the SecYE translocon involved in the protein transport across the inner membrane which, in certain cases, occurs co-translationally (Muller et al., 2001). For YggL, no functional information could be retrieved during this study, whereas the identification of YraL as an rRNA modification enzyme, involved in the $m^4Cm1402$ synthesis at 16S RNA from *E. coli*, has been reported while preparing this work (Kimura & Suzuki, 2009). Therefore, most candidate proteins have some feature that is indicative of a functional relationship with the translation process.

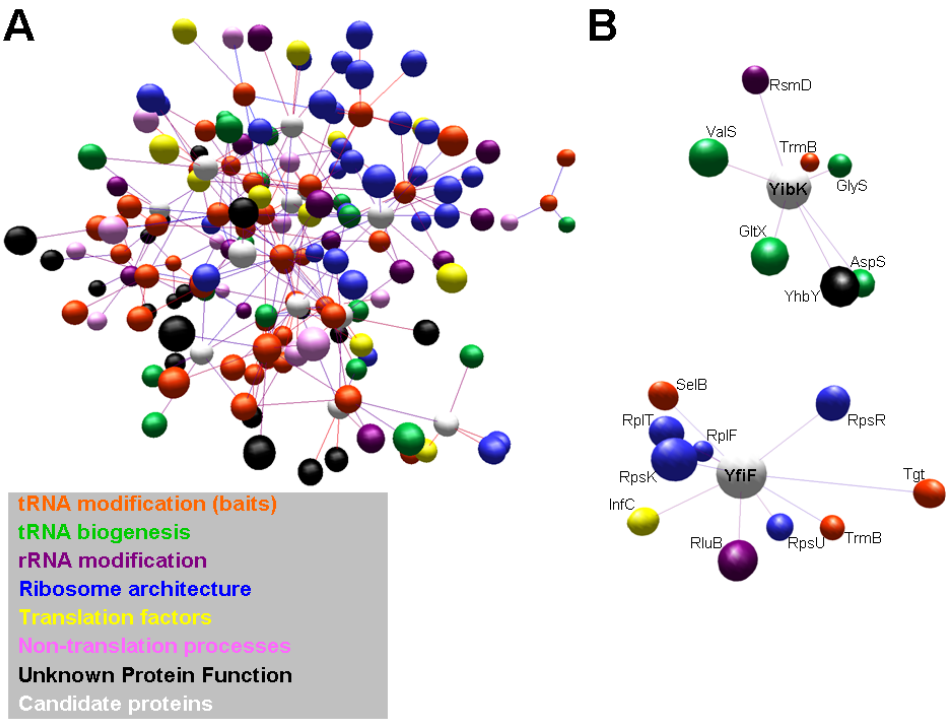

Figure 1. Protein network representing co-evolution-based interactions among the different components of translation machinery. A – Full set of interactions established from tRNA-modifying proteins used as bait in this analysis. B – Local interactions of the YfiF and YibK proteins, respectively.

After the computational analysis, we proceeded to investigate the role in tRNA modification of 11 candidate genes by analysing the stop-codon read-through phenotype and the tRNA modification status conferred by null alleles (*ΔyqcC, ΔyajC, ΔybeZ, ΔybeY, ΔyggL, ΔhflX, ΔychF, ΔyhbC, ΔengC, ΔyibK,* and *ΔyfiF*).

READ-THROUGH PHENOTYPE

The UAG stop codon read-through assay has been extensively used to study how the mutations in translation genes affect the decoding fidelity of ribosomes (Gorini, 1969; Elseviers et al., 1984; Hagervall & Bjork, 1984b; Yim et al., 2003). Therefore, we used P1 transduction to introduce the null alleles of the candidate genes into a wild-type derivative of strain Dev16, which carries the *lacZ105(Amb)* allele (Elseviers et al., 1984; Yim et al., 2006). As a control, we used the strains carrying a null *mnmE* mutation or a *rpsL222* mutation, both of which are well-known to confer a hyperaccuracy

phenotype (Elseviers et al., 1984; Dong & Kurland, 1995; Kurland et al., 1996; Yim et al., 2003; Martinez-Vicente et al., 2005).

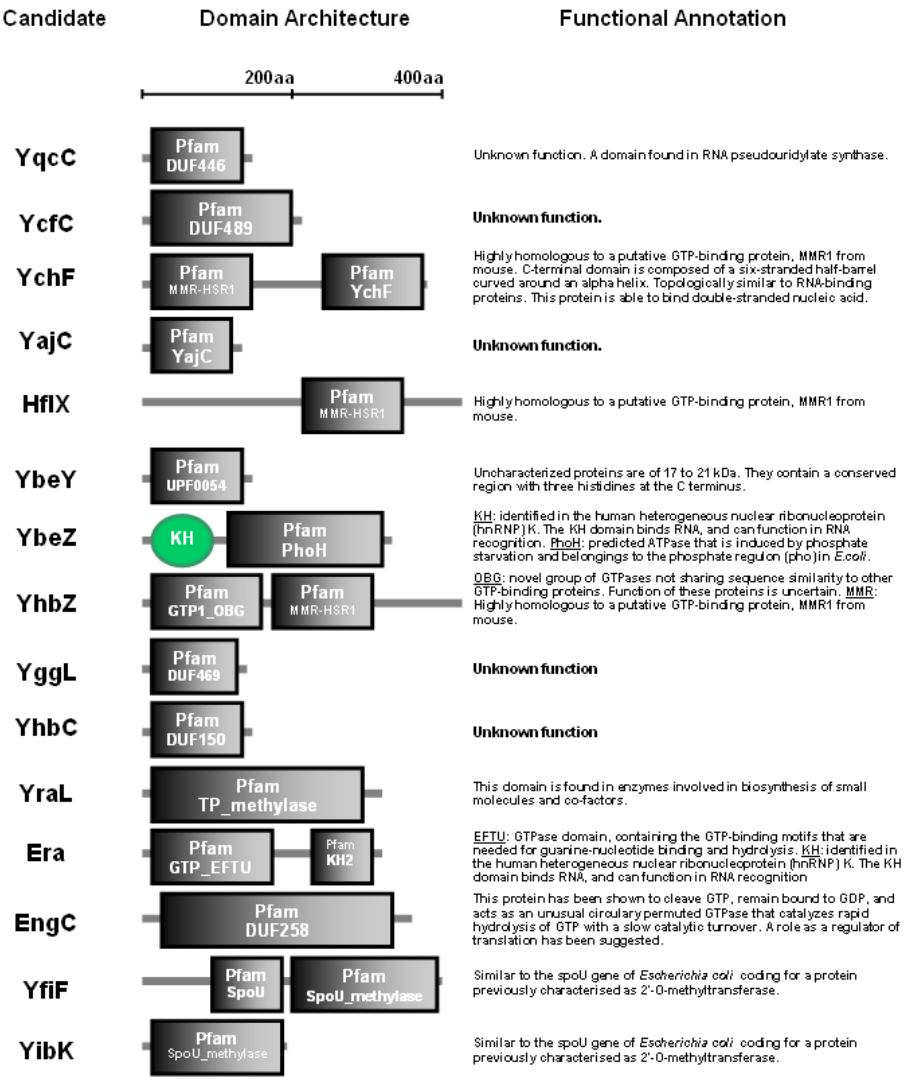

Figure 2. Domain architecture of candidate proteins.

As shown in Figure 3, all the mutants, except *ΔybeY, ΔyibK, and ΔyfiF,* showed significant low read-through levels. The most significant difference was observed in *ΔychF*, which exhibited a 200-fold lower β-galactosidase

activity than that of the wild-type strain, whereas the lowest effect was found in mutant *ΔyhbC*, which still exhibited a significantly lower β-galactosidase activity than the wild-type strain ($p<0.0195$). In general, our results indicate that most of the mutants analyzed could have defects in the translational machinery.

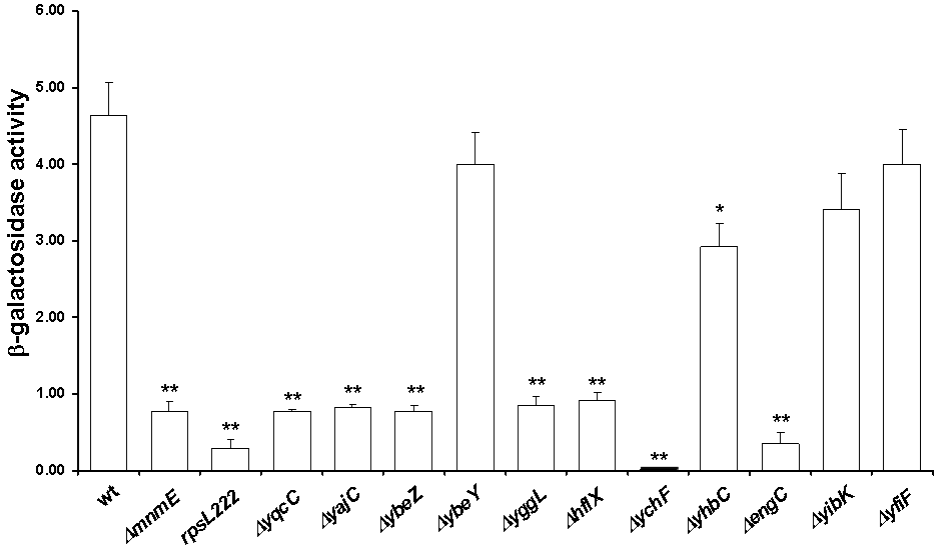

Figure 3. β-galactosidase activity values in the mutants of those genes predicted as tRNA-modifying. The mean is of at least three replicates. * p-value resulting from a t-test and Welch's correction ≤ 0.05; ** p-value ≤ 0.01.

NUCLEOSIDE HPLC PROFILE

The most widespread approach to find evidence of a defect in tRNA modifications consists in comparing the HPLC-nucleoside profiles of hydrolysed tRNAs. In such experiments, defects in nucleoside modifications are evidenced by the fact that their corresponding spectra peaks disappear. This approach has proved most useful for detecting the most apparent modifications. However, the identification of minor tRNA modifications is difficult when complex nucleoside mixtures are analysed. The chromatogram overlapping of both predominant modifications and impurities with minor components seriously limits the detection and reliability of UV absorbance (McCloskey, 1990; Pomerantz & McCloskey, 1990). We compared the HPLC spectra for the tRNA extracted from all the mutants and a wild-type strain, using those peaks corresponding to the most predominant and appreciable

modified nucleosides as references (i.e., Ψ, Q, i^6A, ms^2i^6A, t^6A, m^2A, s^2C, m^7G, Gm, m^1G, hU, s^4U, cmo^5U, and mnm^5s^2U). We could not detect any stable difference between the HPLC chromatogram of each mutant and that corresponding to the wild-type strain (data not shown). This suggests that if the candidate genes were actually tRNA modifying enzymes, then they acted on less apparent modifications.

RNA MASS SPECTROMETRY

Mass spectrometry (MS) to study the modification status of RNAs has become a useful alternative in the last few years (Pomerantz & McCloskey, 1990; Douthwaite & Kirpekar, 2007). This method has been efficiently employed to essentially characterise those genes responsible for rRNA modifications (Purta et al., 2008a; Purta et al., 2008b, 2009). However, some studies on tRNAs have been carried out after isolating specific molecules (Umeda et al., 2005).

The major consideration of the MS approach requires a defined RNA target to be identified, otherwise the data analysis could prove an arduous task. We firstly decided to study the tRNA modification status associated with the impairment of two specific genes, *yfiF* and *yibK*, which according to both the computational results and domain architecture (Figures 1B and S1), seem to be likely involved in RNA modification. Note that the YfiF protein shows interactions with 3 different tRNA-modifying enzymes (Tgt, TrmB, and SelD) and with 7 other members of the translation machinery (Figure 1B), and that YibK shows evolutionary interaction with the tRNA modifying protein TrmB and with five other components of the translation system (Fig. 1B). Additionally, note that although no function can as yet be ascribed to YhbY, which also interacts with YibK, bioinformatic and structural data strongly suggest a RNA binding activity for this protein (Ostheimer et al., 2002). Moreover, proteins YfiF and YibK have a SpoU domain, which has been associated with methylation in RNA. Evolutionary and functional analyses have determined that the methyltransferases carrying such a domain (the SPOUT family) are specifically responsible for synthesising 2'-*O*-methylations (Nm) (Anantharaman et al., 2002b; Purta et al., 2006; Tkaczuk et al., 2007; Purta et al., 2009).

Most of the Nm modifications in *E. coli* tRNA have already been associated with SPOUT methyltransferases. Thus, the TrmH (SpoU) and TrmJ

(YhfQ) proteins were characterised as the enzymes responsible for the Gm18 and Cm32/Um32 methylations in tRNAs, respectively (Persson et al., 1997; Purta et al., 2006). However, the enzymes responsible for the methylation of the 2'-OH group of ribose in the wobble position of $tRNA^{Leu}_{CmAA}$ and $tRNA^{Leu}_{cmnm5UmAA}$ remain to be identified (Tkaczuk et al., 2007). Therefore, YfiF and YibK are attractive candidates to play a role in the modification of both tRNAs (Fig. 4).

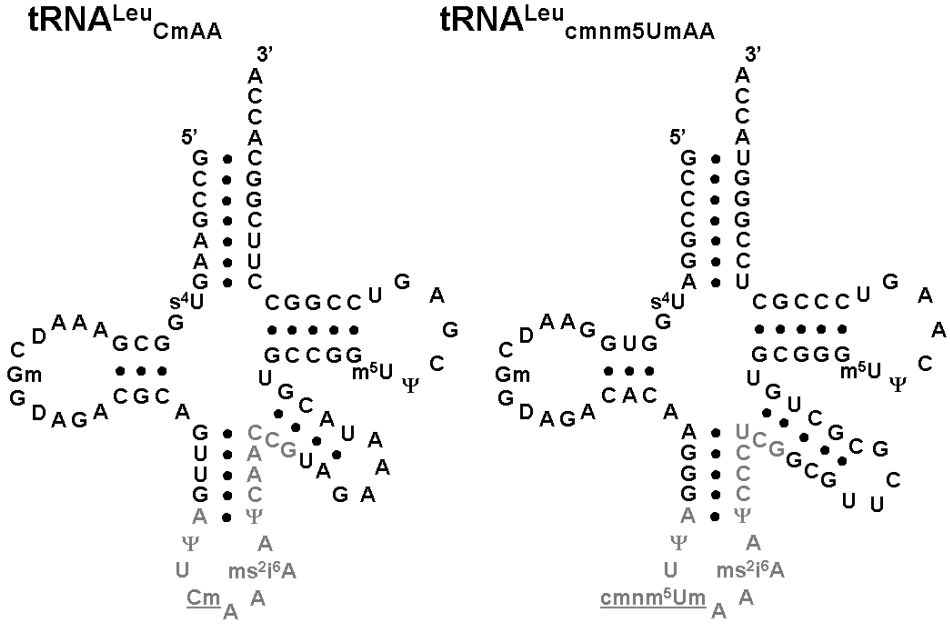

Figure 4. Structure and sequence of the *E.coli* $tRNA^{Leu}UAA$ and $tRNA^{Leu}CAA$. The wobble positions containing the target "Nm" modifications are underlined. Grey nucleotides represent the unique mass fragments expected from the theoretical T1 digestion.

Given our purpose of studying the modification status of the $tRNA^{Leu}_{CmAA}$ and $tRNA^{Leu}_{cmnm5UmAA}$ anticodons in mutants *yfiF* and *yibK*, we designed a challenging approach based on the analysis of the RNase T1 profile of the bulk *E. coli* tRNA. RNase T1 is a guanosine-specific endoribonuclease that yields 2',3'-cyclophosphate ending fragments (Loverix & Steyaert, 2001). We firstly did the *in silico* digestion of the 47 different tRNA sequences found in *Escherichia coli* (Dunin-Horkawicz et al., 2006). After mass correction, which was done by taking into account the modifications present in each fragment (Björk & Hagervall, 2005b), we defined two unique large T1

fragments of 4933.1 Da and 4974.1 Da corresponding to the anticodon region of $tRNA^{Leu}_{CmAA}$ and $tRNA^{Leu}_{cmnm5UmAA}$, respectively.

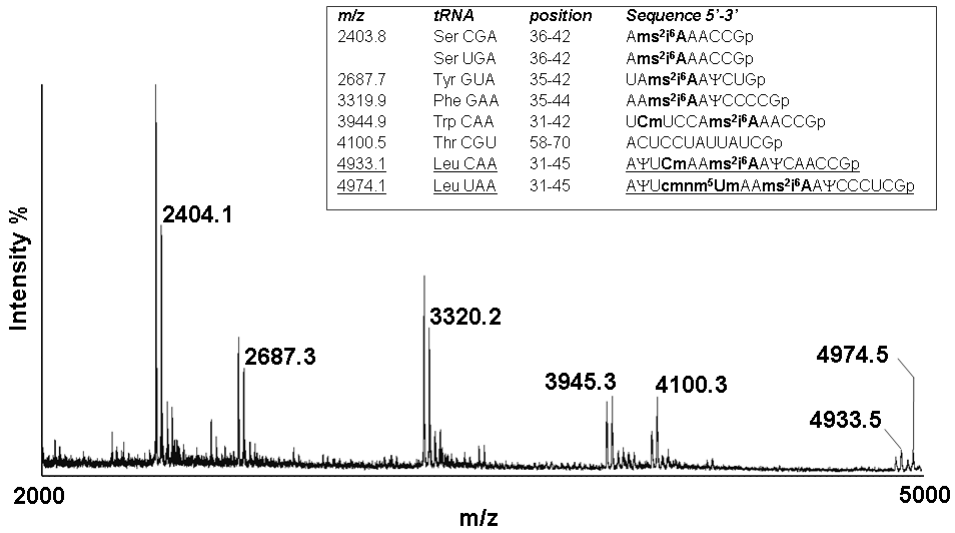

Figure 5. Mass profile from the *E.coli* bulk tRNA digested with RNase T1. Masses of major peaks were matched with the specific T1 fragments obtained from *in silico* digestions. Cyclic forms of peaks are also shown with the respective loss-of-mass corresponding to 18Da.

The T1 digestion of the tRNA pool from a wild-type strain, followed by a reverse chromatography purification of the largest tRNA fragments (see Methods), allowed us to recover a reference MS profile (Figure 5). We could efficiently fit this mass profile to the different fragments obtained from the previous *in silico* T1 digestion. Interestingly, 4 of the 5 major fragments (2404.1, 2687.3, 3320.2, and 3945.3) matched the T1 fragments containing the common ms^2i^6A37 modification. Thus, we detected the T1 fragments originating from $tRNA^{Ser}_{CGA}$ (2403.8 Da), $tRNA^{Ser}_{UGA}$ (2403.8 Da), $tRNA^{Tyr}_{GUA}$ (2687.7), $tRNA^{Phe}_{GAA}$ (3319.9), $tRNA^{Trp}_{CCA}$ (3944.9). Besides, a T1 fragment (4100.3) was assigned to a digestion product originating from $tRNA^{Thr}_{CGU}$ (4100.5), which did not contain the ms^2i^6A37 modification. Despite the $tRNA^{Leu}_{CmAA}$ and $tRNA^{Leu}_{cmnm5UmAA}$ being minor within the *E.coli* tRNA population (Horie et al., 1999), we were able to recover two fragments of 4933.5 and 4974.5 Da that matched the masses predicted for the T1 digestion products of both tRNAs (4933.1 and 4974.1 Da, respectively). Their isolation was probably favoured because they also carried the hydrophobic ms^2i^6A37 modification (Figure 5) which produces fragment retention in reversed-phase chromatography performed on hydrophobic supports, such as the high-

carbon-loading silica (C18) or the poly-styrene-divinylbenzene (POROS Oligos R3) media used in our experiments. We then compared the T1 digestion profile of tRNA from the wild-type strain and the *ΔyfiF* and *ΔyibK* mutants (Figure 6). The fragments corresponding to $tRNA^{Leu}_{CmAA}$ and $tRNA^{Leu}_{cmnm5UmAA}$ were present in the tRNA spectra recovered from *ΔyfiF*, but they exhibited a mass shift in the spectra of *ΔyibK*. Thus, a loss of 14 Da was observed in the specific $tRNA^{Leu}_{CAA}$ and $tRNA^{Leu}_{cmnm5UmAA}$ fragments that matched the lack of the methyl group in the ribose of C34 and U34, respectively (Figure 6).

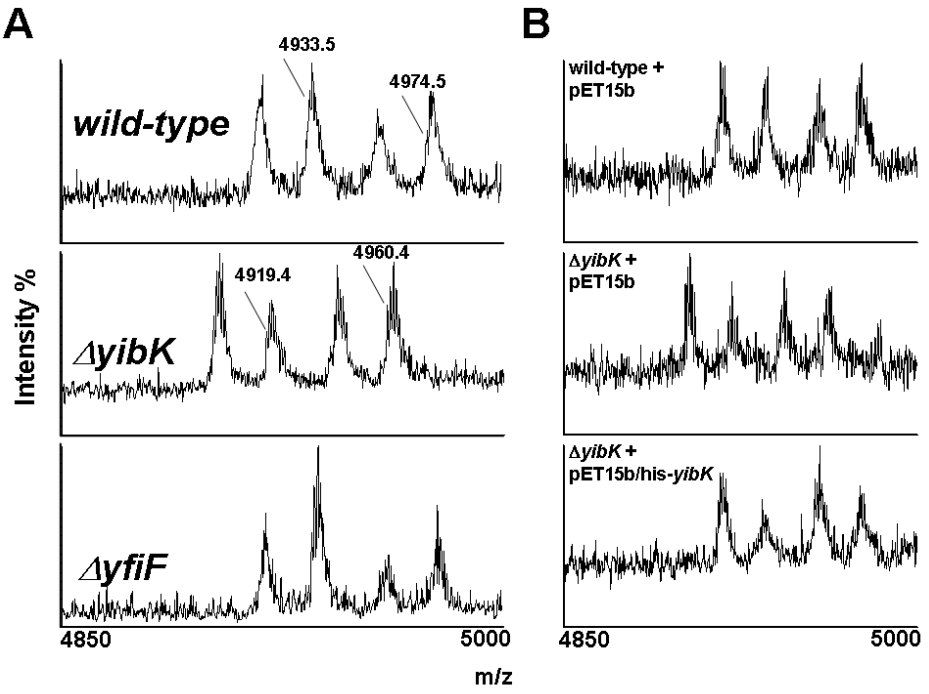

Figure 6. Mass spectra of the T1 fragments from the Leucine tRNA analysed. A - Mass comparison made among the wild-type, *ΔyfiF, and ΔyibK Escherichia coli* strains. Mass values correspond to the monoisotopic mass of the fragment. B – *In vivo* complementation of the tRNA modification in the Leucine tRNAs substrates of YibK.

To confirm that the inactivation of *yibK* is responsible for the spectra change associated with the $tRNA^{Leu}_{CmAA}$ and $tRNA^{Leu}_{cmnm5UmAA}$ anticodon fragments, RNase A was additionally used to digest tRNA, and to thus obtain a new reference MS pattern. RNase A has endonucleolytic activity which is able to hydrolise after pyrimidines, yielding 2',3'-cyclic phosphate ending fragments. The requirement of the 2'-OH group at the ribose ring is

indispensable to complete the transphosphorilation reaction (Burrell, 1993). In this way, nucleosides like Cm or Um (with 2'-*O*-methylations) cannot be recognised for processing, whereas the absence of a modification at the pyrimidine ribose favours RNase A cleavage. Therefore, fragments of 2074.2 and 2162.1 Da, which matched the mass expected for the specific $tRNA^{Leu}_{CmAA}$ and $tRNA^{Leu}_{cmnm5UmAA}$ digestion products, were recovered from the wild-type tRNA, but not from *ΔyibK* tRNA (Figure 7A). The expected RNase A fragments were absent in *ΔyibK* because the lack of ribose methylations in the wobble position of $tRNA^{Leu}_{CmAA}$ and $tRNA^{Leu}_{cmnm5UmAA}$ favours digestion in the unmethylated pyrimidines, which led to the production of a new digestion fragment (AAms^{2}i^{6}AAUp) with a monoisotopic mass of 1755 Da.

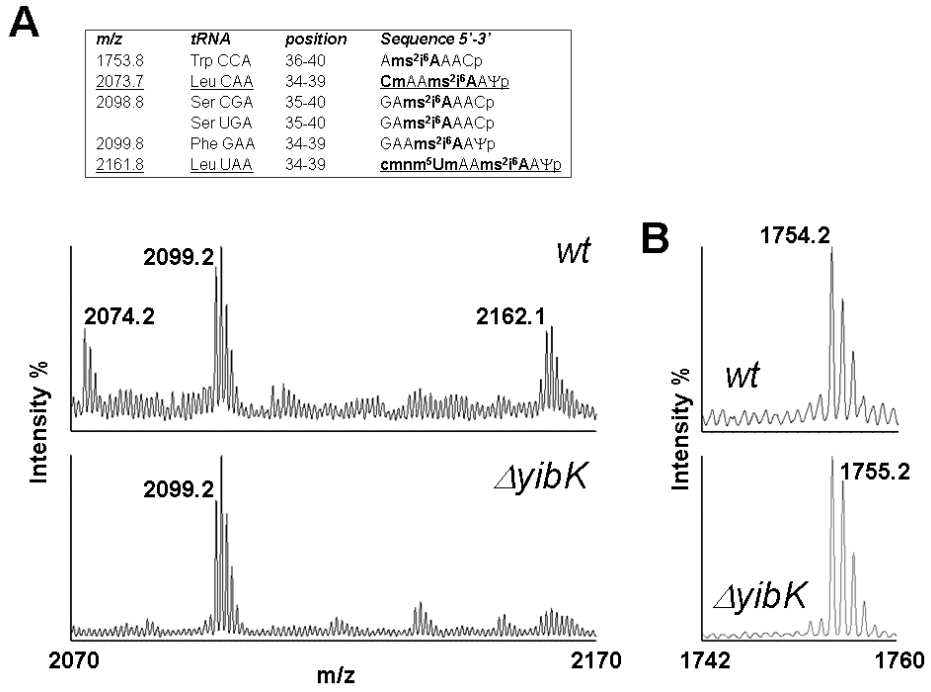

Figure 7. Mass profile from the *E.coli* bulk tRNA digested with RNase A. A – Comparison view between the wild-type and *ΔyibK* strains in the box depicts the major fragments obtained from this profile. B – The zoom in the region of interest where the 2074Da and 2162Da mass fragments are absent in *ΔyibK* beacuse of the loss of the ribose methylation in C34 from $tRNA^{Leu}$CAA and U34 from $tRNA^{Leu}$UAA, respectively.

Figure 7B shows the two peaks of 1754.2 and 1755.2 Da detected in both the wild-type and *yibK* MS spectra. However, it should be noted that the

proportion between both peaks differs in the *yibK* MS spectra due to the increase of the 1755 Da peak. This increase probably results from the contribution of the AAms^{2}i^{6}AAUp fragment from the digestion of the hypomethylated tRNA$^{Leu}_{CmAA}$ and tRNA$^{Leu}_{cmnm5UmAA}$.

The T1 digestion pattern observed in the tRNAs of the wild-type and Δ*yibK* strains (Figure 6A) was used to test the *in vivo* function of *Escherichia coli* YibK. The full-length YibK protein with a histidine tag at the N-terminal was cloned in the pET15b plasmid. The recombinant 6His-YibK protein in mutant *yibK* was expressed by an escape generated from the non-inducible promoters in the plasmid. A mass restoration in both the fragments belonging to tRNA$^{Leu}_{cmnm5UmAA}$ and tRNA$^{Leu}_{CmAA}$ were respectively seen (Figure 6B). In this way, we observe how the recombinant 6His-YibK protein has a complete function under *in vivo* conditions, and could be used to perform *in vitro* reactions in future experiments.

GROWTH RATE AND GROWTH COMPETITION

Mutants *mnmE* and *mnmG* lack a modified U34 in several tRNAs, including tRNA$^{Leu}_{cmnm5UmAA}$, and they grow more slowly than their wild-type counterparts (Yim et al., 2006). Therefore, we investigated whether Um deficiency influences the growth rate. However, no statistical differences between the wild-type and the *ΔyibK* mutant, or between a *ΔmnmG* and a *ΔmnmG/ΔyibK* double mutant, were found when cells were grown in rich medium (Table 3). When *yibK* mutation was combined with the mutation responsible for the modification in ms^{2}i^{6}A37, *miaA*, both present in tRNA$^{Leu}_{CmAA}$ and tRNA$^{Leu}_{cmnm5UmAA}$ (see Fig. 4), no effect was evidenced (Table 3).

Next, we performed a competition experiment which compared not only growth in the logarithmic phase, but also the ability to sustain and recover from the stationary phase. Since the expression of the kanamycin resistance gene (which is present in the *ΔyibK* mutant) involves a cost (Purta et al., 2008b), we compared competition against the wild-type strain (without the kanamycin cassette) of mutants *ΔyibK*, *ΔmnmE*, and *ΔttcA*, all of which carry a kanamycin cassette. We used the last mutant as a control because a *ΔttcA* mutant without a kanamycin cassette has been reported to not show growth disadvantages in competition experiments (Jager et al., 2004). Approximately equal numbers of wild-type and mutant cells (wt *vs.*

ΔyibK::kan, wt *vs. ΔmnmE::kan*, and wt *vs. ΔttcA::kan*) were mixed and incubated during several growth cycles in rich medium (Table 4). The results show that mutants *yibK* and *mnmE* compete more poorly against the wild-type cells than the *ttcA* cells (10-fold lower), which means that the loss of the YibK function involves a biological cost.

Table 3. Growth rate of the *yibK* mutant.

Strain	Doubling time (min)[1]
wild-type	21.0 ± 1.0
ΔyibK	21.4 ± 0.6
ΔmnmG	30.5 ± 1.5
ΔmnmG/ΔyibK	31.5 ± 2.5
ΔmiaA	22.8 ± 0.1
ΔmiaA/ΔyibK	24.2 ± 1.2

[1]Data are presented as the mean of three independent replicates

DISCUSSION

In *Escherichia coli*, 31 different tRNA nucleotide modifications have been reported. However, the identity of those genes which are directly or indirectly involved in the biosynthesis of a significant number of them remains uncertain. In order to identify any additional genes coding for the enzymes involved in tRNA modification, we followed a comparative-genomics approach that exploits the availability of a high number of fully-sequenced bacterial genomes by extracting co-evolutionary information about genes/proteins involved in tRNA-modification. The comparative-genomics based approach used herein has enabled us to detect 15 *E. coli* genes of an unknown function that share a genomic context with known tRNA-modifying enzymes. However, our data analysis indicates that the genomic evolution of these enzymes is tightly linked to the evolution of other proteins related to the translation process. Accordingly, we expected the 15 selected genes that encode for tRNA modification enzymes or other type of proteins to be also involved in either the mRNA decoding process, such as rRNA modification enzymes and translation factors, or the ribosome biogenesis (Fig. 1).

Table 4. Effect of the *yibK* mutation on growth competition

	At the mix time[1]		After six dilutions[1]	
Strain mix (1:1)	**LB**	**LB + kan (ratio)**	**LB**	**LB + kan (ratio)**
wt vs. Δ*ttcA*	2.4×10^8	1.1×10^8 (0.46)	2.1×10^8	9.4×10^5 (0.004)
wt vs. Δ*mnmE*	2.3×10^8	1.1×10^8 (0.48)	1.6×10^8	3.9×10^4 (0.0003)
wt vs. Δ*yibK*	1.9×10^8	1.0×10^8 (0.53)	1.5×10^8	1.8×10^4 (0.0002)

[1]Data are presented as the mean of three independent replicates

To evaluate the quality of our approach, we analysed the stop-codon read-through phenotype and the tRNA modification status conferred by the null mutations in 11 of the 15 genes predicted to be involved in some aspect of the mRNA decoding process. Interestingly, a hyperaccuracy phenotype was observed in eight mutants (Fig. 3), thus supporting a role for the gene affected in translation. The low level UAG-codon read-through observed in the wild-type strain is presumably due to the capability of some specific tRNA to insert an amino acid into the site of the amber mutation in the *lacZ* gene. This tRNA appears to be a substrate for the MnmE·MnmG complex given that the inactivation of this complex by mutations *mnmE* or *mnmG* results in a hyperaccuracy phenotype (Elseviers et al., 1984). The tRNA substrates for MnmE·MnmG are $tRNA^{Lys}_{mnm5s2UUU}$, $tRNA^{Glu}_{mnm5s2UUC}$, $tRNA^{Gln}_{cmnm5s2UUG}$, $tRNA^{Leu}_{cmnm5UmAA}$, and $tRNA^{Arg}_{mnm5UCU}$. The finding that inactivation of *yibK* does not affect the read-through level (Fig. 3) suggests that either $tRNA^{Leu}_{cmnm5UmAA}$ is not the UAG suppressor or that the YibK-dependent modification does not play a role in translational accuracy. Given that the read-through phenotype depends on assay features, the possibility of translational fidelity also being affected in the mutants exhibiting a wild-type phenotype in the read-through assay used in this work cannot be discarded. Regarding *yibK*, its diminished ability to compete with the wild-type strain (Table 4) favours the idea that mRNA decoding is altered in the mutant to some extent. This is not an unexpected consequence if we consider that the modification of 2'-OH to 2'-*O*-methyl is one of the most powerful biological strategies for stabilising the RNA structure (Davis, 1998). The modified nucleoside preferentially adopts the 3'-endo, axial conformation to minimise the steric interactions between the large methyl group and the base, which are more severe for pyrimidines than for purines. It is well established that the modified nucleosides of the anticodon domain of tRNA not only restrict its dynamics, but shape its architecture. This scenario has been proposed to be crucial for ribosomes to overcome the need to constrain or remodel each

tRNA in order to fit the decoding site (Agris, 2008). Thus, the YibK-dependent modification also seems to fulfil this task.

In *E. coli*, many *rpsL* streptomycin-resistant (Sm^R) mutants restrict the natural low-level read-through of some nonsense codons and result in hyperaccurate translation (Kurland et al., 1996). In fact, we used an *rpsL222* mutant as a control in our read-through assays, and found that it exhibits an almost 20-fold reduction of the read-through level if compared to the parental strain. Structural and biochemical studies have identified ribosomal protein S12 as a critical contributor to accuracy by the ribosome and have indicated a critical role for residue Lys43 (Carter et al., 2000; Brodersen et al., 2002; Ogle et al., 2003; Sharma et al., 2007), which changes to Thr in mutant *rpsL222*. In addition, some mutations in the elongator factors have been shown to alter translation fidelity (Kurland et al., 1996). Therefore, according to the hyperaccuracy phenotype conferred by mutations *mnmE* and *rpsL222*, it appears reasonable to propose that the mutants analysed in this work, which also exhibit this phenotype, are affected in those genes involved in the modification of the suppressor tRNA, or in the biogenesis of some ribosomal components or the activity of some translation factors. Considering that the accuracy of the ribosome to read UAG codons depends on multiple factors, we cannot, therefore, pinpoint the specific aspect of the translational process that is affected in most mutants.

The HPLC analysis of the tRNA purified from the 11 mutants did not reveal any differences in relation to the wild-type strain. This suggests that the candidate genes might be involved in less apparent tRNA modifications or in modifications that are barely visible because they are masked by the predominant nucleosides or impurities of processing (McCloskey, 1990; Pomerantz & McCloskey, 1990). Therefore, we used a more sensitive approach based on MALDI-TOF-mass spectrometry to explore the tRNA modification status of some candidate genes. Given that the designed approach requires the mass analysis of a complex mix of tRNA fragments, we chose to firstly study the *yfiF* and *yibK* mutants because the presence of a SPOUT domain in their sequence makes its involvement in RNA modification more plausible. The SPOUT family proteins are involved in the methylation of tRNA and rRNA (Anantharaman et al., 2002b; Tkaczuk et al., 2007). By considering that the enzymes responsible for the ribose methylation occurring in the wobble pyrimidine of $tRNA^{Leu}_{CmAA}$ and

$tRNA^{Leu}_{cmnm5UmAA}$ were unknown, we decided to specifically analyse the mass spectra of those fragments containing the anticodon region of both tRNAs from the wild-type, *yfiF* and *yibK* strains. Whereas no differences were found between the specific fragments from the wild-type and *yfiF* tRNAs, the spectra of the fragments derived from the *yibK* mutant exhibited a mass-shift which could relate with the loss of the ribose methyl group in the wobble position of both tRNAs (Figs. 6-7). Therefore, we conclude that the YibK methyltransferase accomplished the Cm and Um modifications in $tRNA^{Leu}_{CmAA}$ and $tRNA^{Leu}_{cmnm5UmAA}$ respectively, and propose renaming *yibK* as *trmL*, according to the traditional nomenclature for tRNA-modifying genes (Dunin-Horkawicz et al., 2006; Czerwoniec et al., 2009).

Acknowledgements

The author wishes to thank to supervisors of this study: María Eugenia Armengod (CIPF, Valencia, Spain), Toni Galbaldon (CRG, Barcelona, Spain), and Stephen Douthwaite (University of Southern Denmark, Odense, Denmark); to Centro de Investigación Prícipe Felipe for my PhD funding. Thanks also to the EMBO short-term fellowship programme for grant ASTF 186-2009; to Anette Rasmussen and Simon Rose at the Department of Biochemistry and Molecular Biology of the University of Southern Denmark for their invaluable technical assistance in RNA-MS procedures; to Magda Villarroya for her invaluable help. Finally, the author also thanks the National BioResource Project (NIG, Japan): *E. coli,* for providing the knockout mutants used in this study.

Conclusiones

Mediante las diferentes estrategias tanto experimentales como computacionales aplicadas en este estudio, se pudo obtener un gran número de resultados que aportan nuevo conocimiento al campo y objeto principal de esta tesis, la modificación de RNA.

Desde el abordaje experimental se lograron encontrar interesantes datos acerca de la regulación que tiene lugar en algunos genes dedicados a la modificación de RNA. El resolver cómo diversos factores, tanto a nivel transcripcional como a nivel traduccional, afectan la expresión de proteínas se ha convertido en un eje fundamental para explicar interacciones y balance metabólico a través de complejas disciplinas como la biología de sistemas. Los datos presentados en el primer capitulo de esta tesis ofrecen conocimiento acerca de nuevos y cada vez más recurrentes elementos en el ADN que permiten modificar los patrones de expresión de genes. Se pudo confirmar la unidad de transcripción conformada por los genes *mnmG* (*gidA*) y *rsmG* (*gidB*). De igual manera se pudieron establecer nuevos elementos de regulación para *rsmG*, que modulan su exprersión en respuesta al crecimiento de la célula. Junto a ellos se hizo evidente la diferente regulación post-traduccional que presentan ambos enzimas. De esta manera, los elementos que garantizan la expresión autónoma de *rsmG* pueden haber surgido como

respuesta evolutiva para mantener su presencia a niveles adecuados y garantizar la modificación del rRNA, necesaria para el correcto funcionamiento de la maquinaria de traducción.

En cuanto a la caracterización de la proteína RsmG, con ayuda de diferentes estrategias computacionales, y posterior corroboración experimental de sus predicciones, fue posible dilucidar el rol de un amplio número de residuos donde recaen funciones claves como la de unión del cofactor AdoMet, el sutrato y, finalmente, la transferencia química del grupo metil. Actualmente, poco se conoce acerca de los mecanismos mediante los cuales las metiltranferasas realizan metilaciones en el rRNA. Por consiguiente, los datos aportados suponen un buen punto de apoyo para profundizar en el campo.

Finalmente, haciendo uso de información sobre los genomas que han sido secuenciados hasta la fecha, se abordó un complejo estudio para el hallazgo de nuevos genes implicados en la modificación de tRNAs. Este tipo de estrategia, ha generado en los ultimos años importantes descubrimientos en el campo de la modificación de rRNA. No obstante, las enzimas implicadas en tal caso se han resuelto con una mayor facilidad dada la poca variación en cuanto al tipo de modificación que sufre este RNA no codificante. Debido a la gran variedad de las modificaciones en el tRNA, se usó la genómica comparativa para encontrar nuevos genes que participen en este proceso. A partir de las diferentes estrategias utilizadas, se logró obtener un número de posibles genes candidatos que comparten contexto genómico con genes implicados en la modificación de tRNAs. Diferentes estrategias funcionales, de biología evolutiva y de arquitectura de dominios permitieron delimitar el número de genes candidatos con una fuerte predisposición a participar en la modificación de tRNAs. Mediante el ensayo de "*readthorugh*" hemos podido observar la influencia que tienen nuestros candidatos en la traducción de proteínas, confirmando así la implicación de algunos de ellos en este proceso biológico global. Mediante estrategias experimentales con gran sensibilidad en el estudio de la composición de RNAs (MALDI-MS) aplicadas a dos genes candidatos (*yfiF* e *yibK*), seleccionados por la presencia en su secuencia de un dominio SPOUT, pudimos demostrar que uno de ellos, *yibK*, es reponsable de la modificación en 2'-O de la ribosa de la posición de tambaleo del $tRNA^{Leu}_{cmnmU\underline{m}AA}$ y $tRNA^{Leu}_{\underline{Cm}AA}$. Dicha modificación no resulto importante para el crecimiento en fase logarítmica pero es requerida para la salida de las

células de la fase estacionaria de acuerdo con los resultados de competición. Otros genes seleccionados como candidatos son objeto de estudio actualmente en nuestro laboratorio.

References

Agris PF. 2008. Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. *EMBO Rep 9*:629-635.

Agris PF, Vendeix FA, Graham WD. 2007. tRNA's wobble decoding of the genome: 40 years of modification. *J Mol Biol 366*:1-13.

Anantharaman V, Koonin EV, Aravind L. 2001. TRAM, a predicted RNA-binding domain, common to tRNA uracil methylation and adenine thiolation enzymes. *FEMS Microbiol Lett 197*:215-221.

Anantharaman V, Koonin EV, Aravind L. 2002a. Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res 30*:1427-1464.

Anantharaman V, Koonin EV, Aravind L. 2002b. SPOUT: a class of methyltransferases that includes spoU and trmD RNA methylase superfamilies, and novel superfamilies of predicted prokaryotic RNA methylases. *J Mol Microbiol Biotechnol 4*:71-75.

Andersen NM, Douthwaite S. 2006. YebU is a m5C methyltransferase specific for 16 S rRNA nucleotide 1407. *J Mol Biol 359*:777-786.

Andrews JM. 2001. Determination of minimum inhibitory concentrations. *J Antimicrob Chemother 48 Suppl 1*:5-16.

Andrup L, Atlung T, Ogasawara N, Yoshikawa H, Hansen FG. 1988. Interaction of the Bacillus subtilis DnaA-like protein with the Escherichia coli DnaA protein. *J Bacteriol 170*:1333-1338.

Aravind L, Koonin EV. 1999. Novel predicted RNA-binding domains associated with the translation machinery. *J Mol Evol 48*:291-302.

Aravind L, Koonin EV. 2001. THUMP--a predicted RNA-binding domain shared by 4-thiouridine, pseudouridine synthases and RNA methylases. *Trends Biochem Sci 26*:215-217.

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H. 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol 2*:2006 0008.

Bachellier S, Gilson E, Hofnung M, Hill C. 1996. Repeated sequences. In: Neidhardt F, Curtis R, Ingraham J, Lin E, Low B, Magasanik B, Reznikoff W, Riley M, Schaechter M, Umbarger E, eds. *Escherichia coli and Salmonella: cellular and molecular biology*. Washington D.C.: ASM Press. pp 2012-2040.

Basturea GN, Rudd KE, Deutscher MP. 2006. Identification and characterization of RsmE, the founding member of a new RNA base methyltransferase family. *Rna 12*:426-434.

Baumann U, Fischer W, Sprinzl M. 1985. Analysis of modification-dependent structural alterations in the anticodon loop of Escherichia coli tRNAArg and their effects on the translation of MS2 RNA. *Eur J Biochem 152*:645-649.

Bernstein JA, Khodursky AB, Lin PH, Lin-Chao S, Cohen SN. 2002. Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proc Natl Acad Sci U S A 99*:9697-9702.

Bishop AC, Xu J, Johnson RC, Schimmel P, de Crecy-Lagard V. 2002. Identification of the tRNA-dihydrouridine synthase family. *J Biol Chem 277*:25090-25095.

Bjork G, Rasmuson T. 1998. Links between tRNA modification and metabolism and modified nucleosides as tumor markers. In: Grosjean H, Benne R, eds. *Modification and editing of RNA*. Washington D.C.: ASM Press. pp 471-491.

Björk GR, Hagervall TG. 2005a. Chapter 4.6.2, Transfer RNA modification. In: Böck RCI, J. B. Kaper, F. C. Neidhardt, T. Nyström, K. E. Rudd,

and C. L. Squires, ed. *EcoSal—Escherichia coli and Salmonella: cellular and molecular biology*: ASM Press. Washington, D.C. .

Björk GR, Hagervall TG. 2005b. Transfer RNA modification. In: Böck RCI, J. B. Kaper, F. C. Neidhardt, T. Nyström, K. E. Rudd, and C. L. Squires, ed. *EcoSal—Escherichia coli and Salmonella: cellular and molecular biology*: ASM Press. Washington, D.C. .

Bogan JA, Helmstetter CE. 1997. DNA sequestration and transcription in the oriC region of Escherichia coli. *Mol Microbiol 26*:889-896.

Bregeon D, Colot V, Radman M, Taddei F. 2001. Translational misreading: a tRNA modification counteracts a +2 ribosomal frameshift. *Genes Dev 15*:2295-2306.

Bremer H, Dennis P. 1996. Modulation of chemical composition and other parameters of the cell by growth rate. In: Neidhardt F, Curtis R, Ingraham J, Lin E, Low B, Magasanik B, Reznikoff W, Riley M, Schaechter M, Umbarger E, eds. *Escherichia coli and Salmonella: cellular and molecular biology*. Washington D.C.: ASM Press. pp 1553-1570.

Brodersen DE, Clemons WM, Jr., Carter AP, Wimberly BT, Ramakrishnan V. 2002. Crystal structure of the 30 S ribosomal subunit from Thermus thermophilus: structure of the proteins and their interactions with 16 S RNA. *J Mol Biol 316*:725-768.

Bujnicki JM, Oudjama Y, Roovers M, Owczarek S, Caillet J, Droogmans L. 2004. Identification of a bifunctional enzyme MnmC involved in the biosynthesis of a hypermodified uridine in the wobble position of tRNA. *Rna 10*:1236-1242.

Burrell MM. 1993. RNase A (EC 3.1.27.5). *Methods Mol Biol 16*:263-270.

Bystrom AS, Bjork GR. 1982. The structural gene (trmD) for the tRNA(m1G)methyltransferase is part of a four polypeptide operon in Escherichia coli K-12. *Mol Gen Genet 188*:447-454.

Caillet J, Droogmans L. 1988. Molecular cloning of the Escherichia coli miaA gene involved in the formation of delta 2-isopentenyl adenosine in tRNA. *J Bacteriol 170*:4147-4152.

Carter AP, Clemons WM, Brodersen DE, Morgan-Warren RJ, Wimberly BT, Ramakrishnan V. 2000. Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature 407*:340-348.

Ciampi MS, Schmid MB, Roth JR. 1982. Transposon Tn10 provides a promoter for transcription of adjacent sequences. *Proc Natl Acad Sci U S A 79*:5016-5020.

Claros MG. 1995. MitoProt, a Macintosh application for studying mitochondrial proteins. *Comput Appl Biosci 11*:441-447.

Crick FH. 1966. Codon--anticodon pairing: the wobble hypothesis. *J Mol Biol 19*:548-555.

Curran JF. 1998. Modified nucleosides in translation. In: Grosjean H, Benne R, eds. *Modification and Editing RNA*. Washington D.C.: ASM Press. pp 493-516.

Czerwoniec A, Dunin-Horkawicz S, Purta E, Kaminska KH, Kasprzak JM, Bujnicki JM, Grosjean H, Rother K. 2009. MODOMICS: a database of RNA modification pathways. 2008 update. *Nucleic Acids Res 37*:D118-121.

Charette M, Gray MW. 2000. Pseudouridine in RNA: what, where, how, and why. *IUBMB Life 49*:341-351.

Das G, Thotala DK, Kapoor S, Karunanithi S, Thakur SS, Singh NS, Varshney U. 2008. Role of 16S ribosomal RNA methylations in translation initiation in Escherichia coli. *Embo J 27*:840-851.

Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A 97*:6640-6645.

Davis DR. 1998. Biophysical and conformation properties of modified nucleosides in RNA. In: Grosjean H, Benne R, eds. *Modification and Editing RNA*. Washington D.C.: ASM Press. pp 85-102.

De Bie LG, Roovers M, Oudjama Y, Wattiez R, Tricot C, Stalon V, Droogmans L, Bujnicki JM. 2003. The yggH gene of Escherichia coli encodes a tRNA (m7G46) methyltransferase. *J Bacteriol 185*:3238-3243.

Decatur WA, Fournier MJ. 2002. rRNA modifications and ribosome function. *Trends Biochem Sci 27*:344-351.

Del Campo M, Kaya Y, Ofengand J. 2001. Identification and site of action of the remaining four putative pseudouridine synthases in Escherichia coli. *Rna 7*:1603-1615.

Del Campo M, Ofengand J, Malhotra A. 2004. Crystal structure of the catalytic domain of RluD, the only rRNA pseudouridine synthase required for normal growth of Escherichia coli. *Rna 10*:231-239.

Dong H, Kurland CG. 1995. Ribosome mutants with altered accuracy translate with reduced processivity. *J Mol Biol 248*:551-561.

Douthwaite S, Kirpekar F. 2007. Identifying modifications in RNA by MALDI mass spectrometry. *Methods Enzymol 425*:1-20.

Dunin-Horkawicz S, Czerwoniec A, Gajda MJ, Feder M, Grosjean H, Bujnicki JM. 2006. MODOMICS: a database of RNA modification pathways. *Nucleic Acids Res 34*:D145-149.

Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics 5*:113.

El Yacoubi B, Lyons B, Cruz Y, Reddy R, Nordin B, Agnelli F, Williamson JR, Schimmel P, Swairjo MA, de Crecy-Lagard V. 2009. The universal YrdC/Sua5 family is required for the formation of threonylcarbamoyladenosine in tRNA. *Nucleic Acids Res 37*:2894-2909.

Elkins PA, Watts JM, Zalacain M, van Thiel A, Vitazka PR, Redlak M, Andraos-Selim C, Rastinejad F, Holmes WM. 2003. Insights into catalysis by a knotted TrmD tRNA methyltransferase. *J Mol Biol 333*:931-949.

Elseviers D, Petrullo LA, Gallagher PJ. 1984. Novel E. coli mutants deficient in biosynthesis of 5-methylaminomethyl-2-thiouridine. *Nucleic Acids Res 12*:3521-3534.

Elstner M, Andreoli C, Klopstock T, Meitinger T, Prokisch H. 2009. The mitochondrial proteome database: MitoP2. *Methods Enzymol 457*:3-20.

Emilsson V, Kurland CG. 1990. Growth rate dependence of transfer RNA abundance in Escherichia coli. *Embo J 9*:4359-4366.

Esberg B, Leung HC, Tsui HC, Bjork GR, Winkler ME. 1999. Identification of the miaB gene, involved in methylthiolation of isopentenylated A37 derivatives in the tRNA of Salmonella typhimurium and Escherichia coli. *J Bacteriol 181*:7256-7265.

Forsyth RA, Haselbeck RJ, Ohlsen KL, Yamamoto RT, Xu H, Trawick JD, Wall D, Wang L, Brown-Driver V, Froelich JM, C KG, King P, McCarthy M, Malone C, Misiner B, Robbins D, Tan Z, Zhu Zy ZY, Carr G, Mosca DA, Zamudio C, Foulkes JG, Zyskind JW. 2002. A genome-wide strategy for the identification of essential genes in Staphylococcus aureus. *Mol Microbiol 43*:1387-1400.

Gabaldon T. 2008. Comparative genomics-based prediction of protein function. *Methods Mol Biol 439*:387-401.

Gabaldon T, Huynen MA. 2004. Prediction of protein function and pathways in the genome era. *Cell Mol Life Sci 61*:930-944.

Gaur R, Varshney U. 2005. Genetic analysis identifies a function for the queC (ybaX) gene product at an initial step in the queuosine biosynthetic pathway in Escherichia coli. *J Bacteriol 187*:6893-6901.

Gehrke CW, Kuo KC. 1989. Ribonucleoside analysis by reversed-phase high-performance liquid chromatography. *J Chromatogr 471*:3-36.

Golovina AY, Sergiev PV, Golovin AV, Serebryakova MV, Demina I, Govorun VM, Dontsova OA. 2009. The yfiC gene of E. coli encodes an adenine-N6 methyltransferase that specifically modifies A37 of tRNA1Val(cmo5UAC). *Rna 15*:1134-1141.

Gorini L. 1969. The contrasting role of strA and ram gene products in ribosomal functioning. *Cold Spring Harb Symp Quant Biol 34*:101-109.

Green SM, Malik T, Giles IG, Drabble WT. 1996. The purB gene of Escherichia coli K-12 is located in an operon. *Microbiology 142 (Pt 11)*:3219-3230.

Gregory ST, Demirci H, Belardinelli R, Monshupanee T, Gualerzi C, Dahlberg AE, Jogl G. 2009. Structural and functional studies of the Thermus thermophilus 16S rRNA methyltransferase RsmG. *Rna 15*:1693-1704.

Grosjean H. 2009. Nucleic acids are not boring long polymers of only four types of nucleotides: A guided tour. In: Grosjean H, ed. *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution*. Austin, USA: Landes Bioscience. pp 1-18.

Grosjean H, Gaspin C, Marck C, Decatur WA, de Crecy-Lagard V. 2008. RNomics and Modomics in the halophilic archaea Haloferax volcanii: identification of RNA modification genes. *BMC Genomics 9*:470.

Guex N, Peitsch MC. 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis 18*:2714-2723.

Gutgsell N, Englund N, Niu L, Kaya Y, Lane BG, Ofengand J. 2000. Deletion of the Escherichia coli pseudouridine synthase gene truB blocks formation of pseudouridine 55 in tRNA in vivo, does not affect exponential growth, but confers a strong selective disadvantage in competition with wild-type cells. *Rna 6*:1870-1881.

Hagervall TG, Bjork GR. 1984a. Genetic mapping and cloning of the gene (trmC) responsible for the synthesis of tRNA

(mnm5s2U)methyltransferase in Escherichia coli K12. *Mol Gen Genet 196*:201-207.

Hagervall TG, Bjork GR. 1984b. Undermodification in the first position of the anticodon of supG-tRNA reduces translational efficiency. *Mol Gen Genet 196*:194-200.

Hagervall TG, Edmonds CG, McCloskey JA, Bjork GR. 1987. Transfer RNA(5-methylaminomethyl-2-thiouridine)-methyltransferase from Escherichia coli K-12 has two enzymatic activities. *J Biol Chem 262*:8488-8495.

Horie N, Yamaizumi Z, Kuchino Y, Takai K, Goldman E, Miyazawa T, Nishimura S, Yokoyama S. 1999. Modified nucleosides in the first positions of the anticodons of tRNA(Leu)4 and tRNA(Leu)5 from Escherichia coli. *Biochemistry 38*:207-217.

Ikeuchi Y, Kitahara K, Suzuki T. 2008. The RNA acetyltransferase driven by ATP hydrolysis synthesizes N4-acetylcytidine of tRNA anticodon. *Embo J 27*:2194-2203.

Ikeuchi Y, Shigi N, Kato J, Nishimura A, Suzuki T. 2006. Mechanistic insights into sulfur relay by multiple sulfur mediators involved in thiouridine biosynthesis at tRNA wobble positions. *Mol Cell 21*:97-108.

Inoue K, Basu S, Inouye M. 2007. Dissection of 16S rRNA methyltransferase (KsgA) function in Escherichia coli. *J Bacteriol 189*:8510-8518.

Ishihama Y, Schmidt T, Rappsilber J, Mann M, Hartl FU, Kerner MJ, Frishman D. 2008. Protein abundance profiling of the Escherichia coli cytosol. *BMC Genomics 9*:102.

Iwasaki W, Miyatake H, Ebihara A, Miki K. 2004. Crystallization and preliminary X-ray crystallographic studies of the small form of glucose-inhibited division protein A from Thermus thermophilus HB8. *Acta Crystallogr D Biol Crystallogr 60*:515-517.

Jager G, Leipuviene R, Pollard MG, Qian Q, Bjork GR. 2004. The conserved Cys-X1-X2-Cys motif present in the TtcA protein is required for the thiolation of cytidine in position 32 of tRNA from Salmonella enterica serovar Typhimurium. *J Bacteriol 186*:750-757.

Johansen SK, Maus CE, Plikaytis BB, Douthwaite S. 2006. Capreomycin binds across the ribosomal subunit interface using tlyA-encoded 2'-O-methylations in 16S and 23S rRNAs. *Mol Cell 23*:173-182.

Johnston HM, Barnes WM, Chumley FG, Bossi L, Roth JR. 1980. Model for regulation of the histidine operon of Salmonella. *Proc Natl Acad Sci U S A 77*:508-512.

Kambampati R, Lauhon CT. 2003. MnmA and IscS are required for in vitro 2-thiouridine biosynthesis in Escherichia coli. *Biochemistry 42*:1109-1117.

Kammen HO, Marvel CC, Hardy L, Penhoet EE. 1988. Purification, structure, and properties of Escherichia coli tRNA pseudouridine synthase I. *J Biol Chem 263*:2255-2263.

Kanehisa M. 2002. The KEGG database. *Novartis Found Symp 247*:91-101; discussion 101-103, 119-128, 244-152.

Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y. 2008. KEGG for linking genomes to life and the environment. *Nucleic Acids Res 36*:D480-484.

Karita M, Etterbeek ML, Forsyth MH, Tummuru MK, Blaser MJ. 1997. Characterization of Helicobacter pylori dapE and construction of a conditionally lethal dapE mutant. *Infect Immun 65*:4158-4164.

Kaya Y, Del Campo M, Ofengand J, Malhotra A. 2004. Crystal structure of TruD, a novel pseudouridine synthase with a new protein fold. *J Biol Chem 279*:18107-18110.

Kaya Y, Ofengand J. 2003. A novel unanticipated type of pseudouridine synthase with homologs in bacteria, archaea, and eukarya. *Rna 9*:711-721.

Kersey P, Bower L, Morris L, Horne A, Petryszak R, Kanz C, Kanapin A, Das U, Michoud K, Phan I, Gattiker A, Kulikova T, Faruque N, Duggan K, McLaren P, Reimholz B, Duret L, Penel S, Reuter I, Apweiler R. 2005. Integr8 and Genome Reviews: integrated views of complete genomes and proteomes. *Nucleic Acids Res 33*:D297-302.

Kimura S, Suzuki T. 2009. Fine-tuning of the ribosomal decoding center by conserved methyl-modifications in the Escherichia coli 16S rRNA. *Nucleic Acids Res*.

Kimura T, Takagi K, Hirata Y, Hase Y, Muto A, Himeno H. 2008. Ribosome-small-subunit-dependent GTPase interacts with tRNA-binding sites on the ribosome. *J Mol Biol 381*:467-477.

Kirpekar F, Douthwaite S, Roepstorff P. 2000. Mapping posttranscriptional modifications in 5S ribosomal RNA by MALDI mass spectrometry. *Rna 6*:296-306.

Kleckner N, Bender J, Gottesman S. 1991. Uses of transposons with emphasis on Tn10. *Methods Enzymol 204*:139-180.

Kolling R, Gielow A, Seufert W, Kucherer C, Messer W. 1988. AsnC, a multifunctional regulator of genes located around the replication origin of Escherichia coli, oriC. *Mol Gen Genet 212*:99-104.

Kurland C, Hughes D, Ehrenberg M. 1996. Limitations of translational accuracy. In: Neidhardt F, Curtis R, Ingraham J, Lin E, Low B, Magasanik B, Reznikoff W, Riley M, Schaechter M, Umbarger E, eds. *Escherichia coli and Salmonella: cellular and molecular biology*. Washington D.C.: ASM Press. pp 979–1004.

Lamb HK, Thompson P, Elliott C, Charles IG, Richards J, Lockyer M, Watkins N, Nichols C, Stammers DK, Bagshaw CR, Cooper A, Hawkins AR. 2007. Functional analysis of the GTPases EngA and YhbZ encoded by Salmonella typhimurium. *Protein Sci 16*:2391-2402.

Leinfelder W, Forchhammer K, Veprek B, Zehelein E, Bock A. 1990. In vitro synthesis of selenocysteinyl-tRNA(UCA) from seryl-tRNA(UCA): involvement and characterization of the selD gene product. *Proc Natl Acad Sci U S A 87*:543-547.

Leipuviene R, Qian Q, Bjork GR. 2004. Formation of thiolated nucleosides present in tRNA from Salmonella enterica serovar Typhimurium occurs in two principally distinct pathways. *J Bacteriol 186*:758-766.

Lesnyak DV, Sergiev PV, Bogdanov AA, Dontsova OA. 2006. Identification of Escherichia coli m2G methyltransferases: I. the ycbY gene encodes a methyltransferase specific for G2445 of the 23 S rRNA. *J Mol Biol 364*:20-25.

Letunic I, Copley RR, Schmidt S, Ciccarelli FD, Doerks T, Schultz J, Ponting CP, Bork P. 2004. SMART 4.0: towards genomic data integration. *Nucleic Acids Res 32*:D142-144.

Loverix S, Steyaert J. 2001. Deciphering the mechanism of RNase T1. *Methods Enzymol 341*:305-323.

Lu Q, Inouye M. 1998. The gene for 16S rRNA methyltransferase (ksgA) functions as a multicopy suppressor for a cold-sensitive mutant of era, an essential RAS-like GTP-binding protein in Escherichia coli. *J Bacteriol 180*:5243-5246.

Lundgren HK, Bjork GR. 2006. Structural alterations of the cysteine desulfurase IscS of Salmonella enterica serovar Typhimurium reveal substrate specificity of IscS in tRNA thiolation. *J Bacteriol 188*:3052-3062.

Macian F, Perez-Roger I, Armengod ME. 1994. An improved vector system for constructing transcriptional lacZ fusions: analysis of regulation of the dnaA, dnaN, recF and gyrB genes of Escherichia coli. *Gene 145*:17-24.

Martin JL, McMillan FM. 2002. SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Curr Opin Struct Biol 12*:783-793.

Martinez-Vicente M, Yim L, Villarroya M, Mellado M, Perez-Paya E, Bjork GR, Armengod ME. 2005. Effects of mutagenesis in the switch I region and conserved arginines of Escherichia coli MnmE protein, a GTPase involved in tRNA modification. *J Biol Chem 280*:30660-30670.

McCloskey JA. 1990. Constituents of nucleic acids: overview and strategy. *Methods Enzymol 193*:771-781.

Meyer S, Scrima A, Versees W, Wittinghofer A. 2008. Crystal structures of the conserved tRNA-modifying enzyme GidA: implications for its interaction with MnmE and substrate. *J Mol Biol 380*:532-547.

Miller JH. 1990. *A short course in bacterial genetics : a laboratory manual and handbook for Escherichia coli and related bacteria*: Cold Spring Harbor Laboratory Press.

Moukadiri I, Prado S, Piera J, Velazquez-Campoy A, Bjork GR, Armengod ME. 2009. Evolutionarily conserved proteins MnmE and GidA catalyze the formation of two methyluridine derivatives at tRNA wobble positions. *Nucleic Acids Res 37*:7177-7193.

Mueller EG, Buck CJ, Palenchar PM, Barnhart LE, Paulson JL. 1998. Identification of a gene involved in the generation of 4-thiouridine in tRNA. *Nucleic Acids Res 26*:2606-2610.

Muller M, Koch HG, Beck K, Schafer U. 2001. Protein traffic in bacteria: multiple routes from the ribosome to and across the membrane. *Prog Nucleic Acid Res Mol Biol 66*:107-157.

Muramatsu T, Yokoyama S, Horie N, Matsuda A, Ueda T, Yamaizumi Z, Kuchino Y, Nishimura S, Miyazawa T. 1988. A novel lysine-substituted nucleoside in the first position of the anticodon of minor isoleucine tRNA from Escherichia coli. *J Biol Chem 263*:9261-9267.

Murzin AG. 1993. OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *Embo J 12*:861-867.

Nasvall SJ, Chen P, Bjork GR. 2004. The modified wobble nucleoside uridine-5-oxyacetic acid in tRNAPro(cmo5UGG) promotes reading of all four proline codons in vivo. *Rna 10*:1662-1673.

Nasvall SJ, Chen P, Bjork GR. 2007. The wobble hypothesis revisited: uridine-5-oxyacetic acid is critical for reading of G-ending codons. *Rna 13*:2151-2164.

Nilsson K, Lundgren HK, Hagervall TG, Bjork GR. 2002. The cysteine desulfurase IscS is required for synthesis of all five thiolated nucleosides present in tRNA from Salmonella enterica serovar typhimurium. *J Bacteriol 184*:6830-6835.

Nishimura K, Hosaka T, Tokuyama S, Okamoto S, Ochi K. 2007a. Mutations in rsmG, encoding a 16S rRNA methyltransferase, result in low-level streptomycin resistance and antibiotic overproduction in Streptomyces coelicolor A3(2). *J Bacteriol 189*:3876-3883.

Nishimura K, Johansen SK, Inaoka T, Hosaka T, Tokuyama S, Tahara Y, Okamoto S, Kawamura F, Douthwaite S, Ochi K. 2007b. Identification of the RsmG methyltransferase target as 16S rRNA nucleotide G527 and characterization of Bacillus subtilis rsmG mutants. *J Bacteriol 189*:6068-6073.

Nord S, Bylund GO, Lovgren JM, Wikstrom PM. 2009. The RimP protein is important for maturation of the 30S ribosomal subunit. *J Mol Biol 386*:742-753.

Ny T, Bjork GR. 1980. Cloning and restriction mapping of the trmA gene coding for transfer ribonucleic acid (5-methyluridine)-methyltransferase in Escherichia coli K-12. *J Bacteriol 142*:371-379.

O'Connor M, Thomas CL, Zimmermann RA, Dahlberg AE. 1997. Decoding fidelity at the ribosomal A and P sites: influence of mutations in three different regions of the decoding domain in 16S rRNA. *Nucleic Acids Res 25*:1185-1193.

O'Farrell HC, Xu Z, Culver GM, Rife JP. 2008. Sequence and structural evolution of the KsgA/Dim1 methyltransferase family. *BMC Res Notes 1*:108.

Ofengand J, Campo Md. 2005a. Chapter 4.6.1, Modified Nucleosides of Escherichia coli Ribosomal RNA. In: Böck RCI, J. B. Kaper, F. C. Neidhardt, T. Nyström, K. E. Rudd, and C. L. Squires, ed. *EcoSal—Escherichia coli and Salmonella: cellular and molecular biology*: ASM Press. Washington, D.C. .

Ofengand J, Campo Md. 2005b. Modified Nucleosides of Escherichia coli Ribosomal RNA. In: Böck RCI, J. B. Kaper, F. C. Neidhardt, T. Nyström, K. E. Rudd, and C. L. Squires, ed. *EcoSal—Escherichia coli and Salmonella: cellular and molecular biology*: ASM Press. Washington, D.C. .

Ogawa T, Okazaki T. 1991. Concurrent transcription from the gid and mioC promoters activates replication of an Escherichia coli minichromosome. *Mol Gen Genet 230*:193-200.

Ogawa T, Okazaki T. 1994. Cell cycle-dependent transcription from the gid and mioC promoters of Escherichia coli. *J Bacteriol 176*:1609-1615.

Ogle JM, Carter AP, Ramakrishnan V. 2003. Insights into the decoding mechanism from recent ribosome structures. *Trends Biochem Sci 28*:259-266.

Okada N, Noguchi S, Kasai H, Shindo-Okada N, Ohgi T, Goto T, Nishimura S. 1979. Novel mechanism of post-transcriptional modification of tRNA. Insertion of bases of Q precursors into tRNA by a specific tRNA transglycosylase reaction. *J Biol Chem 254*:3067-3073.

Okamoto S, Tamaru A, Nakajima C, Nishimura K, Tanaka Y, Tokuyama S, Suzuki Y, Ochi K. 2007. Loss of a conserved 7-methylguanosine modification in 16S rRNA confers low-level streptomycin resistance in bacteria. *Mol Microbiol 63*:1096-1106.

Osawa T, Inanaga H, Numata T. 2009a. Crystallization and preliminary X-ray diffraction analysis of the tRNA-modification enzyme GidA from Aquifex aeolicus. *Acta Crystallogr Sect F Struct Biol Cryst Commun 65*:508-511.

Osawa T, Ito K, Inanaga H, Nureki O, Tomita K, Numata T. 2009b. Conserved cysteine residues of GidA are essential for biogenesis of 5-carboxymethylaminomethyluridine at tRNA anticodon. *Structure 17*:713-724.

Ostheimer GJ, Barkan A, Matthews BW. 2002. Crystal structure of E. coli YhbY: a representative of a novel class of RNA binding proteins. *Structure 10*:1593-1601.

Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N. 1999. The use of gene clusters to infer functional coupling. *Proc Natl Acad Sci U S A 96*:2896-2901.

Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO. 1999. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci U S A 96*:4285-4288.

Persson BC, Jager G, Gustafsson C. 1997. The spoU gene of Escherichia coli, the fourth gene of the spoT operon, is essential for tRNA (Gm18) 2'-O-methyltransferase activity. *Nucleic Acids Res 25*:4093-4097.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem 25*:1605-1612.

Pomerantz SC, McCloskey JA. 1990. Analysis of RNA hydrolyzates by liquid chromatography-mass spectrometry. *Methods Enzymol 193*:796-824.

Purta E, Kaminska KH, Kasprzak JM, Bujnicki JM, Douthwaite S. 2008a. YbeA is the m3Psi methyltransferase RlmH that targets nucleotide 1915 in 23S rRNA. *Rna 14*:2234-2244.

Purta E, O'Connor M, Bujnicki JM, Douthwaite S. 2008b. YccW is the m5C methyltransferase specific for 23S rRNA nucleotide 1962. *J Mol Biol 383*:641-651.

Purta E, O'Connor M, Bujnicki JM, Douthwaite S. 2009. YgdE is the 2'-O-ribose methyltransferase RlmM specific for nucleotide C2498 in bacterial 23S rRNA. *Mol Microbiol 72*:1147-1158.

Purta E, van Vliet F, Tkaczuk KL, Dunin-Horkawicz S, Mori H, Droogmans L, Bujnicki JM. 2006. The yfhQ gene of Escherichia coli encodes a tRNA:Cm32/Um32 methyltransferase. *BMC Mol Biol 7*:23.

Rasouly A, Schonbrun M, Shenhar Y, Ron EZ. 2009. YbeY, a heat shock protein involved in translation in Escherichia coli. *J Bacteriol 191*:2649-2655.

Reader JS, Metzgar D, Schimmel P, de Crecy-Lagard V. 2004. Identification of four genes necessary for biosynthesis of the modified nucleoside queuosine. *J Biol Chem 279*:6280-6285.

Reese MG. 2001. Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome. *Comput Chem 26*:51-56.

Romanowski MJ, Bonanno JB, Burley SK. 2002. Crystal structure of the Escherichia coli glucose-inhibited division protein B (GidB) reveals a methyltransferase fold. *Proteins 47*:563-567.

Rozenski J, Crain PF, McCloskey JA. 1999. The RNA Modification Database: 1999 update. *Nucleic Acids Res 27*:196-197.

Scrima A, Vetter IR, Armengod ME, Wittinghofer A. 2005. The structure of the TrmE GTP-binding protein and its implications for tRNA modification. *Embo J 24*:23-33.

Schubert HL, Blumenthal RM, Cheng X. 2003. Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem Sci 28*:329-335.

Schultz J, Milpetz F, Bork P, Ponting CP. 1998. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A 95*:5857-5864.

Schuster-Bockler B, Bateman A. 2005. Visualizing profile-profile alignment: pairwise HMM logos. *Bioinformatics 21*:2912-2913.

Sergiev PV, Lesnyak DV, Bogdanov AA, Dontsova OA. 2006. Identification of Escherichia coli m2G methyltransferases: II. The ygjO gene encodes a methyltransferase specific for G1835 of the 23 S rRNA. *J Mol Biol 364*:26-31.

Sergiev PV, Serebryakova MV, Bogdanov AA, Dontsova OA. 2008. The ybiN gene of Escherichia coli encodes adenine-N6 methyltransferase specific for modification of A1618 of 23 S ribosomal RNA, a methylated residue located close to the ribosomal exit tunnel. *J Mol Biol 375*:291-300.

Sha J, Kozlova EV, Fadl AA, Olano JP, Houston CW, Peterson JW, Chopra AK. 2004. Molecular characterization of a glucose-inhibited division gene, gidA, that regulates cytotoxic enterotoxin of Aeromonas hydrophila. *Infect Immun 72*:1084-1095.

Sharma D, Cukras AR, Rogers EJ, Southworth DR, Green R. 2007. Mutational analysis of S12 protein and implications for the accuracy of decoding by the ribosome. *J Mol Biol 374*:1065-1076.

Shields MJ, Fischer JJ, Wieden HJ. 2009. Toward understanding the function of the universally conserved GTPase HflX from Escherichia coli: a kinetic approach. *Biochemistry 48*:10793-10802.

Slany RK, Bosl M, Crain PF, Kersten H. 1993. A new function of S-adenosylmethionine: the ribosyl moiety of AdoMet is the precursor of the cyclopentenediol moiety of the tRNA wobble base queuine. *Biochemistry 32*:7811-7817.

Snel B, Bork P, Huynen M. 2000a. Genome evolution. Gene fusion versus gene fission. *Trends Genet 16*:9-11.

Snel B, Lehmann G, Bork P, Huynen MA. 2000b. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res 28*:3442-3444.

Soma A, Ikeuchi Y, Kanemasa S, Kobayashi K, Ogasawara N, Ote T, Kato J, Watanabe K, Sekine Y, Suzuki T. 2003. An RNA-modifying enzyme

that governs both the codon and amino acid specificities of isoleucine tRNA. *Mol Cell 12*:689-698.

Stern L, Schulman LH. 1977. Role of anticodon bases in aminoacylation of Escherichia coli methionine transfer RNAs. *J Biol Chem 252*:6403-6408.

Stern L, Schulman LH. 1978. The role of the minor base N4-acetylcytidine in the function of the Escherichia coli noninitiator methionine transfer RNA. *J Biol Chem 253*:6132-6139.

Stuart JM, Segal E, Koller D, Kim SK. 2003. A gene-coexpression network for global discovery of conserved genetic modules. *Science 302*:249-255.

Teplyakov A, Obmolova G, Chu SY, Toedt J, Eisenstein E, Howard AJ, Gilliland GL. 2003. Crystal structure of the YchF protein reveals binding sites for GTP and nucleic acid. *J Bacteriol 185*:4031-4037.

Theisen PW, Grimwade JE, Leonard AC, Bogan JA, Helmstetter CE. 1993. Correlation of gene transcription with the time of initiation of chromosome replication in Escherichia coli. *Mol Microbiol 10*:575-584.

Tkaczuk KL, Dunin-Horkawicz S, Purta E, Bujnicki JM. 2007. Structural and evolutionary bioinformatics of the SPOUT superfamily of methyltransferases. *BMC Bioinformatics 8*:73.

Tsuchihashi Z, Brown PO. 1992. Sequence requirements for efficient translational frameshifting in the Escherichia coli dnaX gene and the role of an unstable interaction between tRNA(Lys) and an AAG lysine codon. *Genes Dev 6*:511-519.

Umeda N, Suzuki T, Yukawa M, Ohya Y, Shindo H, Watanabe K, Suzuki T. 2005. Mitochondria-specific RNA-modifying enzymes responsible for the biosynthesis of the wobble base in mitochondrial tRNAs. Implications for the molecular pathogenesis of human mitochondrial diseases. *J Biol Chem 280*:1613-1624.

Urbonavicius J, Skouloubris S, Myllykallio H, Grosjean H. 2005. Identification of a novel gene encoding a flavin-dependent tRNA:m5U methyltransferase in bacteria--evolutionary implications. *Nucleic Acids Res 33*:3955-3964.

van Buul CP, Visser W, van Knippenberg PH. 1984. Increased translational fidelity caused by the antibiotic kasugamycin and ribosomal ambiguity in mutants harbouring the ksgA gene. *FEBS Lett 177*:119-124.

Van Lanen SG, Reader JS, Swairjo MA, de Crecy-Lagard V, Lee B, Iwata-Reuyl D. 2005. From cyclohydrolase to oxidoreductase: discovery of nitrile reductase activity in a common fold. *Proc Natl Acad Sci U S A 102*:4264-4269.

von Mering C, Jensen LJ, Kuhn M, Chaffron S, Doerks T, Kruger B, Snel B, Bork P. 2007. STRING 7--recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res 35*:D358-362.

von Meyenburg K, Jorgensen BB, Nielsen J, Hansen FG. 1982. Promoters of the atp operon coding for the membrane-bound ATP synthase of Escherichia coli mapped by Tn10 insertion mutations. *Mol Gen Genet 188*:240-248.

Walker JE, Gay NJ, Saraste M, Eberle AN. 1984. DNA sequence around the Escherichia coli unc operon. Completion of the sequence of a 17 kilobase segment containing asnA, oriC, unc, glmS and phoS. *Biochem J 224*:799-815.

White DJ, Merod R, Thomasson B, Hartzell PL. 2001. GidA is an FAD-binding protein involved in development of Myxococcus xanthus. *Mol Microbiol 42*:503-517.

Winkler M. 1998. Genetics and regulation of base modification in the tRNA and rRNA of prokaryotes and eukaryotes. In: Grosjean H, Benne R, eds. *Modification and editing of RNA*. Washington D.C.: ASM Press. pp 441-469.

Wolf J, Gerber AP, Keller W. 2002. tadA, an essential tRNA-specific adenosine deaminase from Escherichia coli. *Embo J 21*:3841-3851.

Wolfe MD, Ahmed F, Lacourciere GM, Lauhon CT, Stadtman TC, Larson TJ. 2004. Functional diversity of the rhodanese homology domain: the Escherichia coli ybbB gene encodes a selenophosphate-dependent tRNA 2-selenouridine synthase. *J Biol Chem 279*:1801-1809.

Wu H, Sun L, Blombach F, Brouns SJ, Snijders AP, Lorenzen K, van den Heuvel RH, Heck AJ, Fu S, Li X, Zhang XC, Rao Z, van der Oost J. 2009. Structure of the ribosome associating GTPase HflX. *Proteins 78*:705-713.

Xu Z, O'Farrell HC, Rife JP, Culver GM. 2008. A conserved rRNA methyltransferase regulates ribosome biogenesis. *Nat Struct Mol Biol 15*:534-536.

Yanai I, Derti A, DeLisi C. 2001. Genes linked by fusion events are generally of the same functional category: a systematic analysis of 30 microbial genomes. *Proc Natl Acad Sci U S A 98*:7940-7945.

Yim L, Martinez-Vicente M, Villarroya M, Aguado C, Knecht E, Armengod ME. 2003. The GTPase activity and C-terminal cysteine of the Escherichia coli MnmE protein are essential for its tRNA modifying function. *J Biol Chem 278*:28378-28387.

Yim L, Moukadiri I, Bjork GR, Armengod ME. 2006. Further insights into the tRNA modification process controlled by proteins MnmE and GidA of Escherichia coli. *Nucleic Acids Res 34*:5892-5905.

Zhan C, Fedorov EV, Shi W, Ramagopal UA, Thirumuruhan R, Manjasetty BA, Almo SC, Fiser A, Chance MR, Fedorov AA. 2005. The ybeY protein from Escherichia coli is a metalloprotein. *Acta Crystallogr Sect F Struct Biol Cryst Commun 61*:959-963.

Zheng Y, Roberts RJ, Kasif S. 2002. Genomic functional annotation using co-evolution profiles of gene clusters. *Genome Biol 3*:RESEARCH0060.

Zimmermann RA, Ikeya Y, Sparling PF. 1973. Alteration of ribosomal protein S4 by mutation linked to kasugamycin-resistance in Escherichia coli. *Proc Natl Acad Sci U S A 70*:71-75.

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res 31*:3406-3415.