complex reveal a novel binding mode for TGF-β ligand:receptor interactions. *EMBO J.* 22, 1555–1566
13 Hart, P.J. *et al.* (2002) Crystal structure of the human TβR2 ectodomain–TGF-β3 complex. *Nat. Struct. Biol.* 9, 203–208
14 Kirsch, T. *et al.* (2000) Crystal structure of the BMP-2–BRIA ectodomain complex. *Nat. Struct. Biol.* 7, 492–496
15 Wrana, J.L. *et al.* (1994) Mechanism of activation of the TGF-β receptor. *Nature* 370, 341–347
16 riffith, D.L. *et al.* (1996) Three-dimensional structure of recombinant human osteogenic protein 1: structural paradigm for the transforming growth factor β superfamily. *Proc. Natl. Acad. Sci. U. S. A.* 93, 878–883
17 Grasberger, B. *et al.* (1986) Interaction between proteins localized in membranes. *Proc. Natl. Acad. Sci. U. S. A.* 83, 6258–6262
18 Letzelter, F. *et al.* (1998) The interleukin-4 site-2 epitope determining binding of the common receptor γ chain. *Eur. J. Biochem.* 257, 11–20
19 Schomburg, D. *et al.* (2002) *Springer Handbook of Enzymes*, 2nd edn, Springer Verlag
20 Livnah, O. *et al.* (1999) Crystallographic evidence for preformed dimers of erythropoietin receptor before ligand activation. *Science* 283, 987–990

0968-0004/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tibs.2003.08.001

Genes within genes within bacteria

J. Christopher Ellis and James W. Brown

Department of Microbiology, North Carolina State University, Campus Box 7615, Raleigh, NC 27695-7615, USA

Recently, an unusual gene structure has been described in species of the genus *Thermus*, in which the *rpmH* (ribosomal protein L34) coding sequence was found to be entirely overlapped by the unusually large *rnpA* (RNase P protein subunit) sequence. Gene overlap is common in viruses, but has not been seen to this extent in any bacterium.

Generally, in bacteria, the gene encoding the protein subunit of RNase P (*rnpA*) is located immediately downstream of and in the same orientation as the gene encoding the ribosomal protein L34 (*rpmH*; Figure 1), and the two are located near the origin of replication [1–3]. This co-localization of genes in a wide range of bacterial genomes implies an important linkage in their regulation of expression, but the mechanism of this regulation has not been investigated. These genes in *Escherichia coli* have been demonstrated to be part of the same operon, with two major and one minor promoter upstream of *rpmH*, and two putative transcription-termination signals downstream of *rpmH* [3–5]. The levels of expression of the encoded proteins are quite different; the ribosomal protein

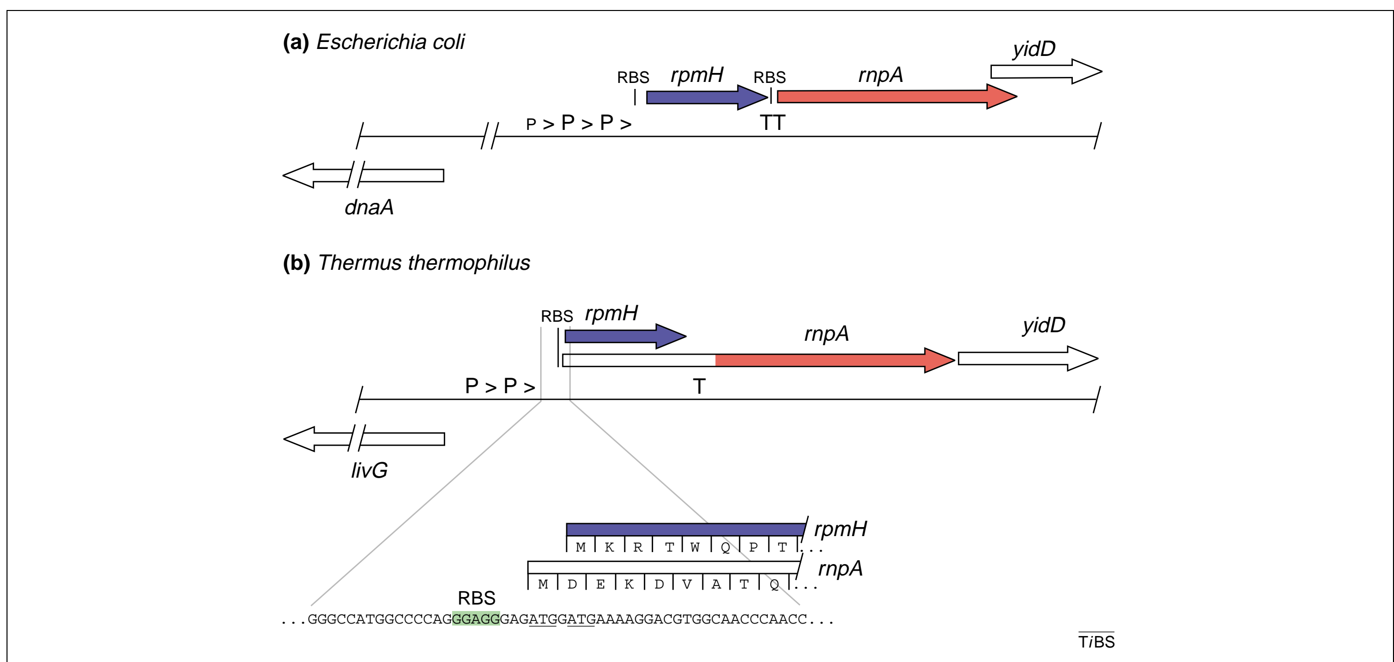

Figure 1. Comparison of *rpmH* with *rnpA* gene structure, which are common in bacteria and *Thermus*. The usual gene structure in bacteria is exemplified by that of *Escherichia coli* (a) [3–5], the overlapping gene structure of *Thermus* species is exemplified by *Thermus thermophilus* (b) [8]. Promoters, putative transcription terminators and ribosome-binding sequences (RBS) for *rpmH* (which encodes the ribosomal protein L34) and *rnpA* (which encodes the protein subunit of RNase P) expression are indicated by P (large for major promoters; small for minor promoters), T and RBS, respectively. Coding sequences are indicated by large arrows; homologous sequences in *rpmH* and *rnpA* are blue and red, respectively. The region of translational initiation in *T. thermophilus* is expanded; the start codons for *rpmH* and *rnpA* are underlined.

Corresponding author: James W. Brown (james_brown@ncsu.edu).

L34 is, of course, abundant in the cell, whereas the RNase P protein is not [6,7]. Protein L34 is produced in greater abundance because two of the three mRNA species transcribed from the operon do not include the RNase P protein coding region [5]. In addition, the mRNA that includes *rnpA* is produced at much lower levels, and the *rnpA* sequence uses codons that are uncommon in *E. coli*, presumably resulting in reduced translation efficiency [3,5].

RpmH and *rnpA* overlap in *Thermus*

A recent paper by Ralph Feltons *et al.* [8] from Roland Hartmann's laboratory describes the unusual overlapping gene structure of *rpmH* and *rnpA* in bacteria of the genus *Thermus* (Figure 1). In *Thermus*, these genes begin with start codons separated by only four base pairs. The second of these start codons initiates the *rpmH* coding sequence, which encodes a typical L34 protein. The upstream start codon initiates the *rnpA* coding sequence in the same orientation, but in the -1 register relative to *rpmH*. The *rnpA* open reading frame continues entirely through *rpmH*, which is some additional distance, and finally includes sequences that are homologous to those of the *rnpA* genes of other bacteria. This overlapping *rnpA* gene results in an unusually long (e.g. 163 amino acids in *Thermus thermophilus*, compared with the usual ~ 120 amino acids), but functional RNase P protein. This overlapping arrangement does not result in compaction of the genes, which is commonly the case in the overlapping genes of viral genomes. The regions of homologous sequences occur in the genome much the same as in other bacteria, only the site of translation initiation of the *rnpA* gene has 'moved' ahead of the upstream gene.

All of the species of *Thermus* investigated had completely overlapping *rnpA* and *rpmH* genes. The genes all started with the same organization of start codons, separated by four base pairs with that of *rnpA* first. Both the *rpmH* and the downstream region of *rnpA* encode typical L34 and RNase P protein sequences, respectively. The length of the *rnpA* sequence varies in different species of *Thermus* in the 'intervening' region, always in multiples of three base pairs, and the amino-acid sequences encoded in this region are quite variable, implying little or no functional constraint. Likewise, the region of *rnpA* sequence that overlaps *rpmH* seems to be conserved with respect to only the L34 encoded amino acids; the RNase P protein sequence encoded in this region is not otherwise conserved.

Implications for gene expression

The differential expression of *rnpA* and *rpmH* in *Thermus* is predicted to result from at least three different aspects of their unusual gene structure. First, the presence of one-to-four potential rho-dependent transcription-termination signals follow the *rpmH* sequence. If the ribosome(s) immediately following the RNA polymerase were translating in the *rpmH* reading frame, this would presumably trigger transcriptional termination at these sites, resulting in an mRNA encoding L34 but not RNase P. Any ribosomes translating in the *rnpA* reading frame of these terminated mRNAs would reach the end of the RNA without encountering a stop codon before the functional part of the RNase P protein is reached; these ribosomes would (conceptually) require the tmRNP (the *ssrA* RNA–*smpB* protein complex) to direct release of the ribosome and target the non-functional truncated RNase P protein for protease degradation [9]. Second, the *rnpA* gene uses unusual codons; these unusual codons are primarily found in the region of overlap with the *rmpH* gene. This implies that the rate of translation for *rnpA* is reduced by ribosomal stalling and/or premature disassociation compared with *rpmH*. Third, the distance of the single ribosome-binding site shared by these two genes is apparently sub-optimally close (three base pairs) to the *rnpA* start codon, but optimally spaced (by seven base pairs) from the *rpmH* start codon. Competition between these start codons should favor translation of L34 over the RNase P protein.

Comparison with MS2 coat protein and lysis protein genes

Although not previously seen in bacteria, gene overlap of this type is common in viruses. A gene arrangement similar to *rpmH* and *rnpA* in *Thermus* can be found in bacteriophage MS2. In this virus, the lysis protein overlaps (in a different reading frame) with the distal portion of the coat cistron and with the proximal portion of the replicase. The N-terminal 40 amino acids of the unusually extended lysis protein are not necessary for functionality and are not well conserved among related phages [10]. It seems that the lysis protein overlap with the coat-protein cistron is not required for additional coding capacity or additional functionality, but couples synthesis of the lysis protein to the coat protein. Therefore, the presence of amino-acid 'extensions' that are not intrinsically functional, but are present to regulate translation of the functional portion of the proteins, are found not only in viruses but also in the chromosomes of some bacteria.

Origin of the gene overlap – how could this happen?

The events generating this unusual gene structure are implied in the structure itself. The fundamental difference between these genes in *Thermus* and other bacteria is the absence of the usual translational start site for the *rnpA*. RNase P is an essential enzyme, and so mutational inactivation of the *rnpA* translational start site would usually be a lethal event. Before the translational start site of *rnpA* could be inactivated in the ancestor of *Thermus*, several conditions must have pre-existed: (i) there must have been a cryptic in-frame translational start site in its modern position upstream of the *rpmH* gene; (ii) there must not have been effective stop codons between this cryptic start site and the original *rnpA* sequence; and (iii) the arbitrary amino-acid-sequence extension must not have disabled the RNase P protein. The most difficult of these would seem to be the absence of stop codons in the *rpmH* gene and the intergenic region in the *rnpA* reading frame. However, the high G+C content of the genomes

of *Thermus* species (60–69%) dramatically reduces the frequency of UAA, UAG and UGA codons between genes and in non-coding reading frames. With these pre-conditions in place, a mutational event inactivating the translational start site of the *rnpA* gene in the ancestor of *Thermus* would have been tolerated, and selective pressure would result in a re-optimization of the sequences based on the new gene structure.

Acknowledgements
Research in the authors' laboratory is supported by NIH grant GM52894 to J.W.B., and is dedicated to the memory of Elizabeth Suzanne Haas (1957–2002).

References

1 Ogasawara, N. and Yoshikawa, H. (1992) Genes and their organization in the replication origin region of the bacterial chromosome. *Mol. Microbiol.* 6, 629–634
2 Salazar, L. *et al.* (1996) Organization of the origins of replication of the chromosomes of *Mycobacterium smegmatis*, *Mycobacterium leprae* and *Mycobacterium tuberculosis* and isolation of a functional origin from *M. smegmatis*. *Mol. Microbiol.* 20, 283–293
3 Hansen, F.G. *et al.* (1985) Physical mapping and nucleotide sequence of the *rnpA* gene that encodes the protein component of ribonuclease P in *Escherichia coli*. *Gene* 38, 85–93
4 Hansen, F.G. *et al.* (1982) The nucleotide sequence of the *dnaA* gene promoter and of the adjacent *rpmH* gene, coding for the ribosomal protein L34, of *Escherichia coli*. *EMBO J.* 1, 1043–1048
5 Panagiotidis, C.A. *et al.* (1992) Modulation of ribonuclease P expression in *Escherichia coli* by polyamines. *Int. J. Biochem.* 24, 1625–1631
6 Dong, H. *et al.* (1996) Growth rate regulation of 4.5 S RNA and M1 RNA the catalytic subunit of *Escherichia coli* RNase P. *J. Mol. Biol.* 261, 303–308
7 Ramagopal, S. (1984) Metabolic changes in ribosomes of *Escherichia coli* during prolonged culture in different media. *Eur. J. Biochem.* 140, 353–361
8 Feltens, R. *et al.* (2003) An unusual mechanism of bacterial gene expression revealed for the RNase P protein of *Thermus* strains. *Proc. Natl. Acad. Sci. U. S. A.* 100, 5724–5729
9 Withey, J.H. and Friedman, D.I. (2002) The biological roles of *trans*-translation. *Curr. Opin. Microbiol.* 5, 154–159
10 Berkhout, B. *et al.* (1985) The amino terminal half of the MS2-coded lysis protein is dispensable for function: implications for our understanding of coding region overlaps. *EMBO J.* 4, 3315–3320

0968-0004/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tibs.2003.08.002

Protein Sequence Motif

POTRA: a conserved domain in the FtsQ family and a class of β-barrel outer membrane proteins

Luis Sánchez-Pulido[1], Damien Devos[1,4], Stéphanie Genevrois[2], Miguel Vicente[3] and Alfonso Valencia[1]

[1]Protein Design Group, Centro Nacional de Biotecnología (CNB-CSIC), Cantoblanco, E-28049 Madrid, Spain
[2]Research Unit in Molecular Biology (URBM), University of Namur (FUNDP), B-5000 Namur, Belgium
[3]Microbial Biotechnology, Centro Nacional de Biotecnología (CNB-CSIC), Cantoblanco, E-28049 Madrid, Spain
[4]University of California, San Francisco, Mission Bay Genentech Hall, 600 16th Street, San Francisco, CA 94143, USA

POTRA (for polypeptide-transport-associated domain) is a novel domain identified in proteins of the ShlB, Toc75, D15 and FtsQ/DivIB families. In most cases, the POTRA domain is associated with a β-barrel outer membrane domain and its function has been experimentally related to polypeptide transport in Toc75 (Tic–Toc protein import system in chloroplast) and ShlB families. In addition to potential key roles in protein transport across the outer membrane and in bacterial septation, the POTRA domain has attractive features for vaccine development in diseases such as cholera, meningitis, gonorrhoea and syphilis.

The 75-kDa subunit of the translocon at the outer envelope of chloroplasts (Toc)-75 is the most abundant protein of the chloroplast outer membrane with a key role in the translocon at the inner envelope of chloroplasts (Tic)–Toc protein import system [1]. *Serratia marcescens* hemolysin lB (ShlB)-related proteins are virulence-factor transporters that are present in the outer membrane of Gram-negative bacteria [2–4]. The D15-related proteins are also found in the outer membrane of Gram-negative bacteria and their physiological function might be related to lipid transport [5] and/or to outer-membrane-protein assembly [6]. Similarity between these families has been proposed based on their transmembrane β-barrel C-terminal regions [7,8]. Furthermore, it has been postulated that Toc75 had its origin in an ancient prokaryotic channel protein of smaller size (ShlB-like family) and evolved by partial gene duplication in the N-terminal region of the protein [7,8].

Here, we characterize these N-terminal regions at the sequence level. We named this region the POTRA domain for polypeptide-transport-associated domain. We offer statistically significant evidence for its presence in the FtsQ/DivIB bacterial division protein family [9] (Figure 1), which is the only case we detected of a POTRA domain not associated with a transmembrane β barrel (Figure 2). This domain has also been found in other less characterized proteins that are associated with a transmembrane β barrel, for example, the YTFM and eukaryotic CGI51 families.

Corresponding author: Luis Sánchez-Pulido (sanchez@cnb.uam.es).