Online journal ISSN - 1676-5680
GMR Genetics and Molecular Research
EVOLUTION AND TECHNOLOGY

Gene expression in *Chromobacterium violaceum*

Rosane Silva[1], Júlia R. Araripe[1], Edson Rondinelli[1,2] and Turán P. Ürményi[1]

[1]Instituto de Biofísica Carlos Chagas Filho,
Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil
[2]Faculdade de Medicina, Universidade Federal do Rio de Janeiro,
Rio de Janeiro, RJ, Brasil
Corresponding author: T.P. Ürményi
E-mail: turmenyi@biof.ufrj.br

Genet. Mol. Res. 3 (1): 64-75 (2004)
Received October 13, 2003
Accepted January 12, 2004
Published March 31, 2004

ABSTRACT. The repertoire of 4,431 open reading frames (ORFs), eight rRNA operons and 98 tRNA genes of *Chromobacterium violaceum* must be expressed in a regulated manner for successful adaptation to a wide variety of environmental conditions. To accomplish this feat, the organism relies on protein machineries involved in transcription, RNA processing and translation. Analysis of the *C. violaceum* genome showed that transcription initiation, elongation and termination are performed by the five well-known RNA polymerase subunits, five categories of sigma 70 factors, one sigma 54 factor, as well as six auxiliary elongation and termination factors. RNA processing is performed by a variety of endonucleases and exonucleases, such as ribonuclease H, ribonuclease E, ribonuclease P, and ribonuclease III, in addition to poly(A) polymerase and specific methyltransferases and pseudouridine synthases. ORFs for all ribosomal proteins, except S22, were found. Only 19 aminoacyl-tRNA synthetases were found, in addition to three aminoacyl-tRNA synthetase-related proteins. Asparaginyl-tRNA^{Asn} is probably obtained by enzymatic modification of a mischarged aminoacyl-tRNA. The translation factors IF-1, IF-2, IF-3, EF-Ts, EF-Tu, EF-G, RF-1, RF-2 and RF-3 are all present in the *C. violaceum* genome, although the absence of selB suggests that *C. violaceum* does not synthesize seleno-

proteins. The components of *trans*-translation, tmRNA and associated proteins, are present in the *C. violaceum* genome. Finally, a large number of ORFs related to regulation of gene expression were also found, which was expected, considering the apparent adaptability of this bacterium.

Key words: *Chromobacterium violaceum*, Gene expression, Gene regulation, Transcription, RNA processing, Translation

INTRODUCTION

The mechanisms of gene expression are essential to make the genetic information contained in the genome explicit, and this genetic information generally codes for proteins. Gene expression mechanisms are also essential so that the correct set of proteins is synthesized at the right time and in adequate amounts, in response to changing conditions inside and/or outside the cell. The cell therefore utilizes a considerable proportion of its biological resources, both genetic and energetic, to ensure accurate protein synthesis to meet current needs. In general, more than 10% of the genes in a genome are devoted to gene expression and its regulation.

Chromobacterium violaceum is a free-living, soil- and water-borne Gram-negative β-proteobacterium found in tropical and subtropical regions (Caldas, 1990). *C. violaceum* has been widely studied, in part due to its biotechnological potential, which includes: a) synthesis of a pigment called violacein with antimicrobial, trypanocidal, antiviral, anticancer, and dermatological properties; b) synthesis of short-chain polyhydroxyalkanoates, which may be used as biodegradable plastics; c) hydrolysis of plastic films, and d) solubilization of gold through a mercury-free process, by which environmental contamination is avoided. The genome of *C. violaceum* has been sequenced and has been found to contain 4,431 open reading frames (ORFs) (Vasconcelos et al., 2003), of which 165 ORFs are devoted to gene expression, excluding regulation. Here we describe the set of genes related to transcription, and RNA processing, translation and regulation of this bacterium.

TRANSCRIPTION

Transcription is the first step in gene expression, and it comprises three well-defined stages, initiation, elongation and termination, each involving different sets of proteins. In *C. violaceum* there are 12 ORFs related to initiation, including RNA polymerase subunits, 11 ORFs related to elongation and 3 ORFs related to termination (Table 1).

Transcription initiation

The multisubunit RNA polymerase (α_2, β, β´, ω), bound to the dissociable initiation sigma factor (σ), provides the majority of the elements for promoter recognition and DNA melting necessary for transcription initiation. The σ, bound to RNA polymerase, recognizes the promoter sequence, in order to initiate transcription. The majority of the transcription initiation that takes place in growing cells is performed by RNA polymerase holoenzyme bound to σ^{70}.

σ^{70} and members of the σ^{70} family direct RNA polymerase to specific promoter elements that are situated 10 and 35 base pairs upstream of transcription initiation sites (Paget and Helmann, 2003). σ^{70} family members may also function as contact points for some activator proteins, such as PhoB and lambda(cl), playing a role in the initiation process itself. The primary sigma factor σ^{70}, which is essential for general transcription in exponentially growing cells, is reversibly associated with RNA polymerase and can be replaced by alternative σ. Alternative σ factors coordinate the expression of genes involved in different functional categories, such as stress responses, morphological development and iron uptake (see below). Sequence comparison of σ^{70} family members reveals conserved regions, which are involved in RNA polymerase binding, promoter recognition and DNA strand separation. As expected, the *C. violaceum* genome contains the genes for σ^{70} (rpoD) and for all RNA polymerase subunits, β (rpoB), β' (rpoC), α (rpoA) and ω (rpoZ).

Transcription elongation and termination

GreA and GreB are involved in the regulation of transcription elongation, including suppression of transcription arrest, enhancement of transcription fidelity, and facilitation of the transition from abortive initiation to productive elongation (Koulich et al., 1998). There are two types of transcription termination in bacteria, Rho-dependent and Rho-independent. Rho-dependent termination requires an RNA/DNA helicase, known as the Rho factor. NusG is a crucial protein that functions as a regulator of Rho-dependent transcription termination, phage lambda N and rRNA transcription antitermination, and phage HK022 Nun termination (Knowlton et al., 2003). Rho-independent transcription termination, where the termination signal appears in the RNA as a hairpin and destabilizes the elongation complex, also plays an essential role in regulating gene expression in prokaryotes. NusA protein controls the hairpin formation promoting termination, and the stabilization of these contacts by phage lambda N protein leads to antitermination (Gusarov and Nuddler, 2001). GreA, GreB, Rho, NusA, NusB and NusG are all present in the *C. violaceum* genome (Table 1).

RNA PROCESSING

RNA processing is a very important step in the production of mature functional RNAs. In prokaryotes, mRNAs, with few exceptions, are the primary transcripts. However, tRNAs and rRNAs undergo considerable processing. ORFs for several proteins related to RNA processing have been found in the *C. violaceum* genome, including poly(A) polymerase, ribonucleases and base modification enzymes (Table 2).

Poly(A) polymerase was discovered in higher eukaryotes and in *E. coli* approximately 40 years ago. The process of polyadenylation of the 3' end of mRNA is nowadays very well known in eukaryotes. Though there are old reports of poly(A) polymerase in *E. coli*, *Bacillus subtilis* and in archea, the study of poly(A) polymerase activities has always been neglected, because only a small fraction of mRNAs could be analyzed. Today we know that the small number of reports does not mean that polyadenylation is rare, but instead it is difficult to detect since the activities of poly(A) polymerase and of nucleases that remove the poly(A) tail keep it to a minimum size (Dreyfus and Regnier, 2002). In eukaryotes, poly(A) tails are the site for binding proteins that stabilize mRNAs, whereas in *E. coli* the poly(A) tail accelerates the deg-

Table 1. ORFs related to transcription.

Function	Gene	Obs.
ATP-dependent helicase	hrpA	
ATP-dependent RNA helicase		
ATP-dependent RNA helicase	rhlE	4 copies
DNA-directed RNA polymerase (alpha subunit)	rpoA	
DNA-directed RNA polymerase (beta subunit)	rpoB	
DNA-directed RNA polymerase (beta subunit)	rpoC	
DNA-directed RNA polymerase (omega subunit)	rpoZ	
N utilization substance protein A	nusA	
N utilization substance protein B	nusB	
RNA helicase	dbpA	
sigma factor 32	rpoH	
sigma factor 38	rpoS	
sigma factor A (sigma 70)	rpoD	
sigma factor E (sigma 24)	rpoE	2 copies
sigma factor for flagellar operon	fliA	2 copies
sigma factor N (sigma 54)	rpoN	
transcription antitermination protein NusG	nusG	
transcription elongation factor GreA	greA	
transcription elongation factor GreB	greB	
transcription termination factor Rho	rho	

radation of RNAs. The presence of poly(A) polymerase in *C. violaceum* as well as in other prokaryotes supports the evolutionary versatility of the function of this RNA modification.

Specific RNA processing reactions produce molecules with unique 5' and 3' ends. Ribonucleases, in particular endonucleases and exonucleases, perform this task. Endonucleases are involved in forming mature RNAs from longer primary transcripts, while exonucleases are involved in trimming the 3' or 5' ends. In addition to their role in maturation processes, ribonucleases also degrade RNA for turnover. The eight rRNA operons and the 98 tRNA genes present in *C. violaceum* are also processed by this repertoire of endonucleases and exonucleases. We found ORFs for ribonuclease H (rnhA and rnhB), ribonuclease E (rne), ribonuclease P (rnpA) and ribonuclease III (rnc). Furthermore, certain bases undergo modification by specific enzymes, conferring the right structure and function to these RNAs. Base modifications of tRNA and rRNA are performed by specific methyltransferases and pseudouridine synthases, such as 16SRNA pseudouridine 516 synthase. This enzyme modifies the nucleotide at position 516 of the 16S rRNA, which is located in the region involved with codon recognition. The *C. violaceum* genome contains several ORFs related to RNA base modifications (Table 2).

TRANSLATION

Protein synthesis is the most conserved mechanism among the cellular processes of genome replication and gene expression. This fact is illustrated by the use of ribosomal RNA as a reference molecule to build phylogenetic trees, which also originally divided living organisms into Bacteria, Archea and Eukarya (Woese et al., 1990). It is not surprising, therefore, that the

Table 2. ORFs related to RNA processing.

Function	Gene	Obs.
Endonucleases and exonucleases		
oligoribonuclease	orn	
poly(A) polymerase	pcnB	
polyribonucleotide nucleotidyltransferase	pnp	
polyribonucleotide nucleotidyltransferase (polynucleotide phosphorylase) (PNPase)	pnp	
ribonuclease BN, putative	?	
ribonuclease E	rne	
ribonuclease HI	rnhA	
ribonuclease HII	rnhB	
ribonuclease II family protein	vacB	
ribonuclease III	rnc	
ribonuclease P protein component	rnpA	
ribonuclease PH	rph	
tRNA nucleotidyltransferase	cca	
tRNA and rRNA base modification		
16S rRNA methyltransferase (sun protein)	sun	2 copies
16S rRNA processing protein rimM	rimM	
16S rRNA pseudouridine synthase A	rluA	
16S rRNA pseudouridylate 516 synthase	rsuA	
23S rRNA pseudouridine synthase C	rluC	2 copies
23S rRNA pseudouridine synthase D	rluD	
dimethyladenosine transferase	ksgA	
pseudouridine synthase	?	
S-adenosylmethionine:tRNA ribosyltransferase-isomerase	queA	
tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase	trmU	
tRNA (guanine-N1)-methyltransferase	trmD	
tRNA (guanosine-2'-O-)-methyltransferase	trmH	3 copies
tRNA (uracil-5-)-methyltransferase	trmA	
tRNA delta(2)-isopentenylpyrophosphate transferase	miaA	
tRNA pseudouridine synthase A	truA	
tRNA pseudouridine synthase B	truB	

set of protein synthesis genes of *C. violaceum* is very similar to that of two other β-proteobacteria, *Neisseria meningitides* (Tettelin et al., 2000) and *Ralstonia solanacearum* (Salanoubat et al., 2002). The translation-related protein-coding genes of *C. violaceum* are shown in Table 3.

The 70S prokaryotic ribosome is composed of a 30S small subunit and a 50S large subunit. Ribosomal protein composition is very conserved among prokaryotes. The 30S subunit may contain 22 ribosomal proteins, S1 to S22, 19 of which (S2 to S20) are obligatory (Lecompte et al., 2002). The 50S subunit may contain 33 ribosomal proteins, L1 to L6, L9 to L25 and L27 to L36. Ribosomal proteins L25 and L30 are the only optional components of the large subunit in some prokaryotes. The *C. violaceum* genome contains the genes for all ribosomal proteins, except S22.

Table 3. ORFs related to translation.

Function	Gene	Obs.
Ribosomal proteins: synthesis and modification		
30S ribosomal protein S1	rpsA	
30S ribosomal protein S2	rpsB	
30S ribosomal protein S3	rpsC	
30S ribosomal protein S4	rpsD	
30S ribosomal protein S5	rpsE	
30S ribosomal protein S6	rpsF	
30S ribosomal protein S7	rpsG	
30S ribosomal protein S8	rpsH	
30S ribosomal protein S9	rpsI	
30S ribosomal protein S10	rpsJ	
30S ribosomal protein S11	rpsK	
30S ribosomal protein S12	rpsL	
30S ribosomal protein S13	rpsM	
30S ribosomal protein S14	rpsN	
30S ribosomal protein S15	rpsO	
30S ribosomal protein S16	rpsP	
30S ribosomal protein S17	rpsQ	
30S ribosomal protein S18	rpsR	
30S ribosomal protein S19	rpsS	
30S ribosomal protein S20	rpsT	
30S ribosomal protein S21	rpsU	
50S ribosomal protein L1	rplA	
50S ribosomal protein L2	rplB	
50S ribosomal protein L3	rplC	
50S ribosomal protein L4	rplD	
50S ribosomal protein L5	rplE	
50S ribosomal protein L6	rplF	
50S ribosomal protein L9	rplI	
50S ribosomal protein L10	rplJ	
50S ribosomal protein L11	rplK	
50S ribosomal protein L12	rplL	
50S ribosomal protein L13	rplM	
50S ribosomal protein L14	rplN	
50S ribosomal protein L15	rplO	
50S ribosomal protein L16	rplP	
50S ribosomal protein L17	rplQ	
50S ribosomal protein L18	rplR	
50S ribosomal protein L19	rplS	
50S ribosomal protein L20	rplT	
50S ribosomal protein L21	rplU	
50S ribosomal protein L22	rplV	
50S ribosomal protein L23	rplW	
50S ribosomal protein L24	rplX	
50S ribosomal protein L25	rplY	

Continued on next page

Table 3. Continued.

Function	Gene	Obs.
50S ribosomal protein L27	rpmA	
50S ribosomal protein L28	rpmB	
50S ribosomal protein L29	rpm	
50S ribosomal protein L30	rpmD	
50S ribosomal protein L31	rpmE	
50S ribosomal protein L32	rpmF	
50S ribosomal protein L33	rpmG	
50S ribosomal protein L34	rpmH	
50S ribosomal protein L35	rpmI	
50S ribosomal protein L36	rpmJ	
ribosomal protein L11 methyltransferase	prmA	
ribosomal-protein-alanine acetyltransferase	rimJ	4 copies
ribosomal-protein-alanine acetyltransferase	rimI	
ribosomal-protein-serine acetyltransferase	rimL	2 copies
Translation factors		
initiation factor IF-1	infA	2 copies
initiation factor IF-2	infB	
initiation factor IF-3	infC	
elongation factor G (EF-G)	fusA	
elongation factor P (EF-P)	efp	
elongation factor Ts (EF-Ts)	tsf	
elongation factor Tu	tufB	
elongation factor Tu	tufA	
peptide chain release factor 1	prfA	
peptide chain release factor 2	prfB	authentic conserved frameshift
peptide chain release factor 3	prfC	
ribosomal binding factor A	rbfA	
ribosome recycling factor	frr	
tRNA aminoacylation		
alanyl-tRNA synthetase	alaS	2 copies (with an ala-tRNA synthetase-related protein)
arginyl-tRNA synthetase	argS	
aspartyl-tRNA synthetase	aspS	
cysteinyl-tRNA synthetase	cysS	
glutaminyl-tRNA synthetase	glnS	
glutamyl-tRNA synthetase	gltX	2 copies (with a glu-tRNA synthetase-related protein)
glycyl-tRNA synthetase, alpha	glyQ	
glycyl-tRNA synthetase, beta	glyS	
histidyl-tRNA synthetase	hisS	2 copies (with a his-tRNA synthetase-related protein)
isoleucyl-tRNA synthetase	ileS	

Continued on next page

Table 3. Continued.

Function	Gene	Obs.
leucyl-tRNA synthetase	leuS	
lysyl-tRNA synthetase	lysS	
methionyl-tRNA synthetase	metG	
phenylalanyl-tRNA synthetase, alpha	pheS	
phenylalanyl-tRNA synthetase, beta	pheT	
prolyl-tRNA synthetase	proS	
seryl-tRNA synthetase	serS	
threonyl-tRNA synthetase	thrS	
tryptophanyl-tRNA synthetase	trpS	2 copies
tyrosyl-tRNA synthetase	tyrS	
valyl-tRNA synthetase	valS	
methionyl-tRNA formyltransferase	fmt	
glutamyl-tRNA (Gln) amidotransferase, subunit A	gatA	
glutamyl-tRNA (Gln) amidotransferase, subunit B	gatB	
glutamyl-tRNA (Gln) amidotransferase, subunit C	gatC	
peptidyl-tRNA hydrolase	pth	
queuine-tRNA ribosyltransferase	tgt	

Aminoacyl-tRNA synthesis provides the interface between nucleotide sequence and the amino acid sequence, and therefore is crucial for translation accuracy. The attachment of the correct amino acid to a specific tRNA is carried out by aminoacyl-tRNA synthetases in a two-step process of ATP-dependent activation, followed by transfer of the amino acid to the tRNA (Ibba and Söll, 2000). This process involves preferential binding of the correct amino acid and editing of the incorrect one, and is capable of discriminating very similar amino acids. Most organisms, like *E. coli*, possess 20 aminoacyl-transferases, one for each of the 20 amino acids (Ibba and Söll, 2001). Many prokaryotes, however, synthesize some of their aminoacyl-tRNAs by enzymatic modification of a mischarged amino acid. Some archea synthesize in this way up to 5 of the 21 aminoacyl-tRNAs needed, such as formylmethionyl-tRNA^{fMet}, glutaminyl-tRNA^{Gln}, asparaginyl-tRNA^{Asp} and selenocysteinyl-tRNA^{Sec} (Blanquet et al., 2000).

C. violaceum contains the genes of all aminoacyl-tRNA synthetases, except asparaginyl-tRNA^{Asp}. This observation is in agreement with what was previously described for other β-proteobacteria, such as *Neisseria* spp (Ibba and Söll, 2001). Asparaginyl-tRNA^{Asn} is probably obtained through enzymatic modification of a mischarged aspartyl-tRNA^{Asn} by a glutamyl-tRNA^{Gln} amidotransferase, the genes of which (gatA, gatB and gatC) were also found in the *C. violaceum* genome. It has been found that glutamyl-tRNA^{Gln} amidotransferases can also function as aspartyl-tRNA^{Asn} amidotransferases (Salazar et al., 2001). In fact, synthesis of asparaginyl-tRNA^{Asn} by amidotransferase reactions is so widespread in bacteria that it has been speculated that asparagine was the last amino acid recruited for protein synthesis (Ibba and Söll, 2001). In addition, three aminoacyl-tRNA synthetase-related protein genes were found, which in other organisms have been shown to have functions unrelated to aminoacylation (Schimmel and De Pouplana, 2000); they are glutamyl-tRNA synthetase-related protein, histidyl-tRNA synthetase-related protein and alanyl-tRNA synthetase-related protein. The function of these genes in *C. violaceum* is unclear.

Translation is divided into three steps, initiation, elongation and termination, which are controlled by specific factors. These factors, IF-1, IF-2, IF-3, EF-Ts, EF-Tu, EF-G, RF-1, RF-2 and RF-3, are all present in the *C. violaceum* genome, including the ribosome recycling factor. Some proteins require a modified amino acid, selenocysteine, which is incorporated at the time of peptide chain elongation (Bock, 2000). There is no codon for this modified amino acid, which is inserted at an in-frame UGA termination codon recognized by a selenocysteinyl-$tRNA^{Sec}$-specific translation factor called SelB. The selB gene was not found, suggesting that *C. violaceum* cannot synthesize selenoproteins.

tmRNA is an abundant, small RNA, universally present in bacteria (Williams, 2002), which has a dual nature, having characteristics of both tRNA and mRNA (Williams and Bartel, 1996). It was first described in *E. coli* as being encoded by the ssrA gene, and is also called 10Sa RNA or SsrA RNA (Ray and Apirion, 1979; Chauhan and Apirion, 1989; Oh et al., 1990). tmRNA plays, together with SmpB (small protein B), a small basic protein that binds specifically to it, a role in the process of *trans*-translation. When ribosomes reach the 3' end of an mRNA that lacks a termination codon, they become stalled. *Trans*-translation is the mechanism whose function is to recycle the stalled ribosomes, adding a C-terminal peptide tag to the unfinished protein, targeting it for proteolysis (Tu et al., 1995; Keiler et al., 1996; Vioque and de la Cruz, 2003). *Trans*-translation also comes into play when bacteria are placed in sub-lethal concentrations of protein synthesis inhibitors (Vioque and de la Cruz, 2003).

The tmRNA of *C. violaceum* has already been described in strains ATCC07461 and ATCC12472 by Felden et al. (2001) and by Williams (2002), respectively. There are two different sequence versions that are available at the tmRNA Web site (http://www.indiana.edu/~tmrna) and the three differing nucleotides possibly reflect intraspecies variation. The SmpB protein gene is also present in the *C. violaceum* genome. Other proteins are described to associate less tightly with tmRNA: ribosomal protein S1, RNase R (a member of the RNase II family), PrsA (an enzyme required for *de novo* synthesis of nucleotides, tryptophan and histidine) and SAF (a possible formyltransferase) (Karzai and Sauer, 2001). Both the ribosomal protein S1 and a member of the RNase II family are also present in the *C. violaceum* genome. In general, tmRNA is not essential for cell viability under standard lab conditions, but charged tmRNA seems to be essential for *Neisseria gonorrhoeae* viability (Huang et al., 2000). It would be interesting to know if charged tmRNA is also essential to *C. violaceum*.

TRANSCRIPTIONAL REGULATION

Bacteria have one housekeeping σ and a variable number of alternative σ which possess different promoter-recognition properties. A repertoire of different σ can be used by the cell to modify the transcription program in response to stress.

Free-living microorganisms are exposed to a series of variable environmental conditions, such as alterations in the abundance of nutrients, changes in temperature and toxic compounds. These environmental variations require fast adaptive responses, usually triggering transcriptional activation of specific genes. Alternative σ are specialized in controlling some regulons that are activated by specific stress conditions, growth transitions, and morphological changes. In addition, transcription is modulated by activators and repressors, most of which alter the expression of transcription of the housekeeping σ.

The σ^{70} family has been divided, based on similarity, into four major groups. Group 1

includes σ^{70} and its orthologs, such as RpoD, responsible for the transcription of housekeeping genes. Group 2 members, such as σ^{S} (RpoS), the general stress regulator of *E. coli* and the most studied member of this group, are not necessary for growth. Group 3 members are more divergent in sequence, and can be divided, with related proteins, into groups with similar functions, such as heat shock (σ^{32} or RpoH), flagellar biosynthesis (σ^{F} or FliA) and sporulation. Group 4 members, originally called extracytoplasmic function (Ecf) family members, such as σ^{E} (RpoE) and FecI, have reduced sequence similarity to the other σ^{70} groups. The superscript numbers on σ indicate the molecular weight. *Escherichia coli* has a pool of six alternative σ, such as the σ^{70} family, σ^{70} (RpoD), σ^{S} (RpoS), σ^{32} (RpoH), σ^{F} (FliA), σ^{E} (RpoE), and FecI. In *C. violaceum*, ORFs corresponding to σ^{70} (RpoD), σ^{S} or σ^{38} (RpoS), σ^{32} (RpoH), σ^{F} or σ^{28} (FliA) and σ^{E} or σ^{24} (RpoE) were found.

Most alternative σ are related in sequence and structure to σ^{70}. However, many bacteria also have a second distinct type of σ called σ^{N} (also called σ^{54} or RpoN), which shares no sequence similarity with the σ^{70} family. σ^{54} family members utilize an enhancer sequence and ATP to form an open complex, while σ^{70} members do not (Studholme and Buck, 2000). Although most bacteria contain multiple members of the σ^{70} family, they usually have no more than one representative of the σ^{54} family. Different from other alternative σ^{70}, the σ^{54} or σ^{N} is correlated with the expression of gene products that have a broad range of functions. The diverse types of functions of genes that are under RNA polymerase and σ^{N} control also vary within groups of bacteria; this situation does not suggest any clear set of specific functions. Functions described to be under σ^{54} control are nitrogen assimilation, fixation and related regulatory functions, glutamine transport, nodulation, phage shock, and others (Gruberl and Gross, 2003). Moreover, the distribution of σ^{54} family members varies among microorganisms. It is present in *N. meningitidis*, *Yersinia pestis*, *Salmonella typhi*, and absent in *Haemophilus influenzae*. The genome of *C. violaceum* contains σ^{54} (rpoN), although the set of functions regulated by this protein is unknown.

Anti-sigma factors modulate the regulons controlled by some σ at the level of transcription, by inhibitory binding. The broad range of cell processes regulated by anti-sigma factors includes bacteriophage growth, sporulation, stress response, flagellar biosynthesis, pigment production, ion transport, and virulence (Helmann, 1999). We found an ORF that potentially codes for anti-sigma 28 (FlgM) in the *C. violaceum* genome.

Transcriptional activators and repressors belong to a variety of well-defined families of transcription factors, which interact with alternative sigma factors involved in bacterial cell response to stressors. More than 60% of the ORFs related to transcriptional regulators identified in *C. violaceum* show high similarity scores (e value of at least -17) to transcription factors described for *E. coli*. One hundred and seventy ORFs were identified as transcription factors that bind DNA through helix-turn-helix motifs and were tentatively classified into families. The largest group of transcription regulators found in *C. violaceum* appears to belong to the LysR and AraC families, as has also been described in other bacteria (Schell, 1993; Martin and Rosner, 2001). An interesting feature of LysR transcriptional regulators is that some of them behave as sensors of physiological changes and can bind directly to different kinds of molecules. The AraC family is known by their homology to a 99-amino acid segment described for the first transcriptional activator in *E. coli* B (Martin and Rosner, 2001). In *Ralstonia solanacearum*, which can be found free-living in soil, a large number of transcription factor genes have also been reported (Salanoubat et al., 2002). The large number of transcriptional regulators found in

the *C. violaceum* genome suggests considerable adaptive capability in patterns of gene expression in response to changes in the environment.

CONCLUDING REMARKS

The genome sequencing of *C. violaceum*, a bacterium found in tropical and subtropical regions, has provided a large amount of information about how this free-living organism can adapt to different environments. Similarity searches and comparison analyses of the ORFs related to gene expression and regulation identified in this organism showed the expected conservation in these systems. The protein components of transcription, RNA processing, translation, and regulation are very similar to that of other members of β-proteobacteria, such as *R. solanacearum* and *Neisseria* spp., and show a considerable versatility in gene expression. This preliminary analysis may be used as the starting point for further studies to improve our understanding of the gene expression mechanisms of *C. violaceum*.

ACKNOWLEDGMENTS

This study was undertaken within the context of the Brazilian National Genome Program, a consortium funded in December 2000 by the Ministério da Ciência e Tecnologia (MCT) through the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

REFERENCES

Blanquet, S., Mechulam, Y. and **Schmitt, E.** (2000). The many routes of bacterial transfer RNAs after aminoacylation. *Curr. Opin. Struct. Biol. 10*: 95-101.

Bock, A. (2000). Biosynthesis of selenoproteins - an overview. *Biofactors 11*: 77-78.

Caldas, L.R. (1990). Um pigmento nas águas negras. *Cienc. Hoje 11*: 56-67.

Chauhan, A.K. and **Apirion, D.** (1989). The gene for a small stable RNA (10Sa RNA) of *Escherichia coli*. *Mol. Microbiol. 3*: 1481-1485.

Dreyfus, M. and **Regnier, P.** (2002). The poly(A) tail of mRNAs: bodyguard in eukaryotes. *Cell 111*: 611-613.

Felden, B., Massire, C., Westhof, E., Atkins, J.F. and **Gesteland, R.F.** (2001). Phylogenetic analysis of tmRNA genes within a bacterial subgroup reveals a specific structural signature. *Nucleic Acids Res. 29*: 1602-1607.

Gruberl, T.M. and **Gross, C.A.** (2003). Sigma subunit and the partitioning of bacterial transcription space. *Annu. Rev. Microbiol. 57*: 441-466.

Gusarov, I. and **Nudler, E.** (2001). Control of intrinsic transcription termination by N and NusA: the basic mechanisms. *Cell 107*: 437-449.

Helmann, J.D. (1999). Anti-sigma factors. *Curr. Opin. Microbiol. 2*: 135-141.

Huang, C., Wolfgang, M.C., Withey, J., Koomey, M. and **Friedman, D.I.** (2000). Charged tmRNA but not tmRNA-mediated proteolysis is essential for *Neisseria gonorrhoeae* viability. *EMBO J. 19*: 1098-1107.

Ibba, M. and **Söll, D.** (2000). Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem. 69*: 617-650.

Ibba, M. and **Söll, D.** (2001). The renaissance of aminoacyl-tRNA synthesis. *EMBO Reports 21*: 382-387.

Karzai, A.W. and **Sauer, R.T.** (2001). Protein factors associated with the SsrA-SmpB tagging and ribosome rescue complex. *Proc. Natl. Acad. Sci. USA 98*: 3040-3044.

Keiler, K.C., Waller, P.R.H. and **Sauer, R.T.** (1996). Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science 271*: 990-993.

Knowlton, J.R., Bubunenko, M., Andrykovitch, M., Guo, W., Routzahn, K.M., Waugh, D.S., Court, D.L. and **Ji, X.** (2003). A spring-loaded state of NusG in its functional cycle is suggested by X-ray crystallography and supported by site-directed mutants. *Biochemistry 42*: 2275-2281.

Koulich, D., Nikiforov, V. and **Borukhov, S.** (1998). Distinct functions of N- and C-terminal domains of GreA, an *Escherichia coli* transcript cleavage factor. *Mol. Biol. 276*: 379-389.

Lecompte, O., Ripp, R., Thierry, J.-C., Moras, D. and **Poch, O.** (2002). Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Res. 30*: 5382-5390.

Martin, R.G. and **Rosner, J.L.** (2001). The AraC transcriptional activators. *Curr. Op. Microbiol. 4*: 132-137.

Oh, B.K., Chauhan, A.K., Isono, K. and **Apirion, D.** (1990). Location of a gene (ssrA) for a small, stable RNA (10Sa RNA) in the *Escherichia coli* chromosome. *J. Bacteriol. 172*: 4708-4709.

Paget, M.S. and **Helmann, J.D.** (2003). The sigma70 family of sigma factors. *Genome Biol. 4*: 203.

Ray, B.K. and **Apirion, D.** (1979). Characterization of 10S RNA: a new stable RNA molecule from *Escherichia coli*. *Mol. Gen. Genet. 174*: 25-32.

Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S., Ariat, M., Billaut, A., Brottlier, P., Camus, J.C., Cattolico, L., Chandler, M., Cholsne, N., Claudel-Renard, C., Cunnac, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schlex, T., Siguler, P., Thébault, P., Whalen, M., Wincker, P., Levy, M., Weissenbach, J. and **Boucher, C.A.** (2002). Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature 415*: 497-501.

Salazar, J.C., Zúñiga, R., Raczniak, G., Becker, H., Söll, D. and **Orellana, O.** (2001). A dual-specific Glu-tRNA^{Gln} and Asp-tRNA^{Asn} amidotransferase is involved in decoding glutamine and asparagine codons in *Acidithiobacillus ferroxidans*. *FEBS Lett. 500*: 129-131.

Schell, M.A. (1993). Molecular biology of the LysR family of transcriptional regulators. *Annu. Rev. Microbiol. 47*: 597-626.

Schimmel, P. and **De Pouplana, L.R.** (2000). Footprints of aminoacyl-tRNA synthetases are everywhere. *Trends Biochem. Sci. 25*: 207-209.

Studholme, D.J. and **Buck, M.** (2000). The biology of enhancer-dependent transcriptional regulation in bacteria: insights from genome sequences. *FEMS Microbiol. Lett. 186*: 1-9.

Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C., Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F., Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D., Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D., Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E., Cittone, H., Clark, E.B., Cotton, M.D., Utterback, T.R., Khouri, H., Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Masignani, V., Pizza, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R., Rappuoli, R. and **Venter, J.C.** (2000). Complete genome sequence of *Neisseria meningitides* serogroup B strain MC58. *Science 287*: 1809-1815.

Tu, G.-F., Reid, G.E., Zhang, J.-G., Moritz, R.L. and **Simpson, R.J.** (1995). C-teminal extension of truncated recombinant proteins in *Escherichia coli* with a 10as RNA decapeptide. *J. Biol. Chem. 270*: 9322-9326.

Vasconcelos, A.T.R., Almeida, D.F., Hungria, M., Guimarães, C.T., Antônio, R.V., Almeida, F.C., Almeida, L.G.P., Almeida, R., Alves-Gomes, J.A., Andrade, E.M., Araripe, J., Araújo, M.F.F., Astolfi-Filho, S., Azevedo, V., Baptista, A.J., Bataus, L.A.M., Batista, J.S., Beló, A., van den Berg, C., Bogo, M., Bonatto, S., Bordignon, J., Brigido, M.M., Brito, C.A., Brocchi, M., Burity, H.A., Camargo, A.A., Cardoso, D.D.P., Carneiro, N.P., Carraro, D.M., Carvalho, C.M.B., Cascardo, J.C.M., Cavada, B.S., Chueire, L.M.O., Creczynski-Pasa, T.B., Fagundes, N., Falcão, C.L., Fantinatti, F., Farias, I.P., Felipe, M.S.S., Ferrari, L.P., Ferro, J.A., Ferro, M.I.T., Franco, G.R., Freitas, N.S.A., Furlan, L.R., Gazzinelli, R.T., Gomes, E.A., Gonçalves, P.R., Grangeiro, T.B., Grattapaglia, D., Grisard, E.C., Hanna, E.S., Jardim, S.N., Laurino, J., Leoi, L.C.T., Lima, L.F.A., Loureiro, M.F., Lyra, M.C.C.P., Madeira, H.M.F., Manfio, G.P., Maranhão, A.Q., Martins, W.S., Mauro, S.M.Z., Medeiros, S.R.B., Meissner, R.V., Moreira, M.A.M., Nascimento, F.F., Nicolás, M.F., Oliveira, J.G., Oliveira, S.C., Paixão, R.F.C., Parente, J.A., Pedrosa, F.O., Pena, S.D.J., Pereira, J.O., Pereira, M., Pinto, L.S.R.C., Pinto, L.S., Porto, J.I.R., Potrich, D.P., Ramalho-Neto, C.E., Reis, A.M.M., Rigo, L.U., Rondinelli, E., Santos, E.B.P., Santos, F.R., Schneider, M.P.C., Seuanez, H.N., Silva, A.M.R., Silva, A.L.C., Silva, D.W., Silva, R., Simões, I.C., Simon, D., Soares, C.M.A., Soares, R.B.A., Souza, E.M., Souza, K.R.L., Souza, R.C., Steffens, M.B.R., Steindel, M., Teixeira, S.R., Ürményi, T.P., Vettore, A., Wassem, R., Zaha, A. and **Simpson, A.J.G.** (2003). The complete genome sequence of *Chromobacterium violaceum* reveals remarkable and exploitable bacterial adaptability. *Proc. Natl. Acad. Sci. USA 100*: 11660-11665.

Vioque, A. and **de la Cruz, J.** (2003). *Trans*-translation and protein synthesis inhibitors. *FEMS Microb. Lett. 218*: 9-14.

Williams, K.P. (2002). The tmRNA Website: invasion by an intron. *Nucleic Acids Res. 30*: 179-182.

Williams, K.P. and **Bartel, D.P.** (1996). Phylogenetic analysis of tmRNA secondary structure. *RNA 2*: 1306-1310.

Woese, C.R., Kandler, O. and **Wheelis, M.L.** (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. USA 87*: 4576-4579.