Nucleic Acids Research, 1998, Vol. 26, No. 23 **5379–5387**

In vitro and *in vivo* secondary structure probing of the *thrS* leader in *Bacillus subtilis*

Dong Luo, Ciarán Condon, Marianne Grunberg-Manago and Harald Putzer*

UPR 9073, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, 75005 Paris, France

Received August 10, 1998; Revised and Accepted October 13, 1998

ABSTRACT

The *Bacillus subtilis thrS* gene is a member of the T-box gene family in Gram-positive organisms whose expression is regulated by a tRNA-mediated transcriptional antitermination mechanism involving a direct tRNA:mRNA interaction. The complex leader sequences of these genes share only short stretches of primary sequence homology, but a common secondary structure has been proposed by comparing the leaders of many genes of this family. The proposed mechanism for the tRNA:mRNA interaction depends heavily on the secondary structure model, but is so far only supported by genetic evidence. We have studied the structure of the *B.subtilis thrS* leader in solution, in protection experiments using both chemical and enzymatic probes. The *thrS* leader structure was also probed *in vivo* using dimethylsulphate and the *in vitro* and *in vivo* data are in good accordance. We have organized the *thrS* leader into three major domains comprising six separate stem–loops. All but one of the short sequences conserved in this gene family are present in loop structures. The ACC specifier codon proposed to interact with the tRNA$^{Thr}_{GGU}$ isoacceptor is present in a bulge and probably exists in a stacking conformation. The proposed antiterminator structure is not visible in transcripts containing the terminator, but was probed using a transcript with the 3′-half of the terminator deleted and its folding appears consistent with the regulatory model. The leader sequences, and in particular the specifier domains, of the other genes of this family can be folded similarly to the experimentally solved *thrS* structure.

INTRODUCTION

The *Bacillus subtilis* threonyl-tRNA synthetase gene (*thrS*) is a member of a family of Gram-positive genes which are controlled by tRNA-mediated transcriptional antitermination in response to starvation for a particular amino acid (1–6). All of these genes have a highly structured untranslated leader mRNA of ~300 nt in length, culminating in a factor-independent transcription terminator, just upstream of the translation initiation site (2,7–9). Upstream of the leader terminator is a highly conserved sequence of ~14 nt, called the T-box, part of which can base pair with a conserved complementary sequence in the 5′-proximal half of the terminator stem, to form an alternative and mutually exclusive antiterminator structure (1). Following starvation for the cognate amino acid, the resulting uncharged tRNA is thought to interact with the leader region in two places to stabilize the antiterminator (Fig. 1). The first of these interactions is with a triplet, the 'specifier codon', encoding the regulatory amino acid in a classical codon:anticodon base pairing (1). This regulatory codon appears to be bulged out of a putative stem–loop structure that occurs early in the leader region of genes regulated in this manner (1,3,4). The second interaction is thought to occur by base pairing between the 5′-UGGN′-3′ sequence of the T-box and the complementary NCCA-3′ end of the uncharged tRNA (3,10). This latter interaction is thought to stabilize the antiterminator conformation, allowing RNA polymerase to continue transcription of the structural gene. An additional level of complexity was introduced with the discovery of an RNA cleavage site immediately upstream of the terminator. In the case of *thrS*, cleavage at this site is specifically increased under threonine starvation conditions (11). The mRNA is processed by a functional analogue of *Escherichia coli* RNase E and the increased stability of the cleaved RNA is thought to significantly contribute to the final level of induction (12).

In the genome sequence of *B.subtilis* (13) a total of 19 genes can be identified which appear to have the leader elements required for this type of regulation. In addition to 14 genes coding for aminoacyl-tRNA synthetases, there are three operons or genes involved in amino acid biosynthesis, one putative amino acid transporter and one gene of unknown function. Up to now, all evidence supporting this regulatory mechanism has come from genetic experiments. Since the secondary structure of the leader mRNA is critical to the proposed mechanism of regulation, we decided to determine the structure of the *thrS* leader both *in vitro* and *in vivo* to see whether it fits with the current model. Our data also permit us to position several short conserved sequence stretches which are crucial for regulation (C.Condon and H.Putzer, unpublished results; 14), but for which no obvious function has been found.

MATERIALS AND METHODS

Bacterial strains and growth conditions

Echerichia coli strain JM109 [*recA1, endA1, gyrA86, thi, hsdR17, supE44, relA1*, λ$^-$, Δ(*lac-proAB*) F′(*traD36, proAB, lacI*^{q}, *lacZ*, ΔM15)] was used for plasmid amplifications and genetic

*To whom correspondence should be addressed. Tel: +33 1 43298226; Fax: +33 1 40468331; Email: putzer@ibpc.fr

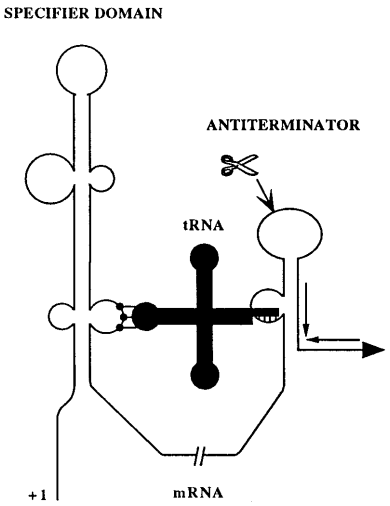

Figure 1. Model for the regulation of expression of genes by tRNA-mediated antitermination. The schematic depicts the genetically identified points of interaction between the mRNA leader and the corresponding tRNA which is believed to favour the formation of the antiterminator structure. Filled black circles represent the specifier codon. The scissors indicate a cleavage site upstream of the terminator believed to occur in many genes regulated by this mechanism. Inverted arrows represent the terminator structure.

manipulations. *Bacillus subtilis* strain HP48 was used in the *in vivo* DMS modification experiments. It corresponds to strain HP18 harbouring the *thrS* leader overproducing plasmid pHMS23 (15). *Escherichia coli* strains were grown in Luria–Bertani (LB) broth and *B.subtilis* strains in M9 minimal medium supplemented with trace elements (16). Threonine starvation was achieved by the addition of 600 µg/ml DL-threonine hydroxamate to a minimal medium culture at OD_{600} of ~0.4. Cells were harvested 20 min after the addition of threonine hydroxamate. Arginine starvation was achieved in the same way by the addition of 3 µg/ml arginine hydroxamate for 20 min.

Plasmid constructs

pHMS50. A 0.4 kb fragment was amplified by PCR using two primers, HP128 (5′-AGAATTCTAATACGACTCACTATAGG-GAGATTAAGAAAGACACACG-3′) and HP129 (5′-TATCTA-GATATCTTCTGTTGTTGTTCC-3′), from the plasmid template pHMS7 (2). Oligonucleotide HP128 contains an *Eco*RI cloning site, a T7 RNA polymerase promoter and the first 21 nt of the *thrS* leader transcript. Oligonucleotide HP129 is complementary to nt 364–384 of the *thrS* mRNA and contains an *Xba*I cloning site at its 5′-end. The new plasmid permits synthesis of the *thrS* leader transcript *in vitro* from its native +1 position.

pHMS23. This plasmid overproduces the *thrS* leader region from a replicative plasmid and has been described previously (15).

In vitro transcription by T7 RNA polymerase

In general, 1 µg of template (plasmid pHMS50 digested by *Eco*RV or *Eco*NI) was used in the T7 RNA polymerase *in vitro* transcription reaction. The reaction mixture contained 40 mM Tris–HCl, pH 7.5, 6 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl, 10 mM DTT, 100 U RNase inhibitor RNasin, 0.5 mM each ATP, CTP, GTP and UTP and 20 U T7 RNA polymerase. All enzymes are from Promega. The reaction was carried out at 37°C for 90 min. DNA templates were then digested by RNase-free RQ1 DNase from Promega and the transcripts were purified on a NAP-10 column (Pharmacia).

Chemical and enzymatic probing *in vitro*

Chemical probing. In general, 0.1 µg of transcript synthesized *in vitro* was denatured by heating for 1 min at 80°C followed by cooling immediately on ice for 1 min. Transcripts were renatured in AN (50 mM Na cacodylate, pH 7.5, 5 mM $MgCl_2$, 60 mM KCl) or AS buffer (50 mM Na cacodylate, pH 7.5, 1 mM EDTA) for kethoxal and DMS modification and in BN (50 mM Na borate, pH 8.0, 5 mM $MgCl_2$, 60 mM KCl) or BS buffer (50 mM Na borate, pH 8.0, 1 mM EDTA) for CMCT modification for 15 min at room temperature. The renatured RNAs were then incubated with either kethoxal (0.25 or 0.5 mg/ml) or DMS (diluted 1/640 or 1/1280 in 20% ethanol) for 5 min at 30°C or with CMCT (5 or 10 mg/ml) for 10 min at 30°C in the presence of 2 µg of carrier 16S rRNA from *Lactococcus lactis*. The control reactions were done in the absence of chemical reagents under the corresponding conditions. The reactions were stopped by addition of 25 mM Na borate, pH 7.0, for kethoxal or by ethanol precipitation. The chemical modifications were analysed by primer extension.

Enzyme probing. Aliquots of 0.1 µg of transcripts synthesized *in vitro* were denatured and renatured in AN buffer as described for the chemical modifications. The renatured transcripts were digested with RNase T1 (Boehringer Mannheim) at 2×10^{-3} or 4×10^{-3} U/µl or with RNase V1 (Pharmacia) at 5×10^{-4} or 10^{-3} U/µl for 15 min at 30°C in the presence of 2 µg carrier RNA. RNase CL3 (Boehringer Mannheim) digestions were performed in the same buffer at 10^{-4} or 2×10^{-4} U/µl for 20 min at 30°C. Enzyme cleavages were analysed by primer extension.

Primer hybridization and extension with reverse transcriptase. Three oligonucleotides were used to analyse the chemical and enzymatic modifications of the *thrS* leader transcript: HP9 (5′-CCC-GTCAGCCATTACTGTTCC-3′), which is complementary to nt 145–165 of the leader, HP27 (5′-CCTTGACTGCTCCATCAG-GAAATG-3′), which is complementary to nt 329–352 in the coding sequence of the *thrS* mRNA, and HP106 (5′-AAAGGGA-CGAGGTTGAAT-3′), which corresponds to nt 304–321 in the terminator. HP106 was used only for the 'antiterminator' transcript. Aliquots of 50–100 nmol of RNA were hybridized with 1–3 pmol of 5′-labeled primer in 5 µl of distilled water as described (17). Primer extension was performed at 45°C for 30 min with 1 U AMV reverse transcriptase. Reactions were stopped with formamide containing sample buffer and one half of the total reaction was loaded on 6 or 8% polyacrylamide electrophoresis gels.

DMS probing *in vivo*

Strain HP48 was grown in M9 minimal medium in the presence (starvation) or absence (normal growth) of threonine hydroxamate or arginine hydroxamate. RNA was modified *in vivo* according to a previously described protocol (18) and the modifications were revealed by primer extension using oligonucleotides HP9 (above) and HP241 (5′-CGGAATCGTGGTTCCACCC-3′), complementary to nt 223–241 of the *thrS* mRNA immediately upstream of the major processing site at residue 241.

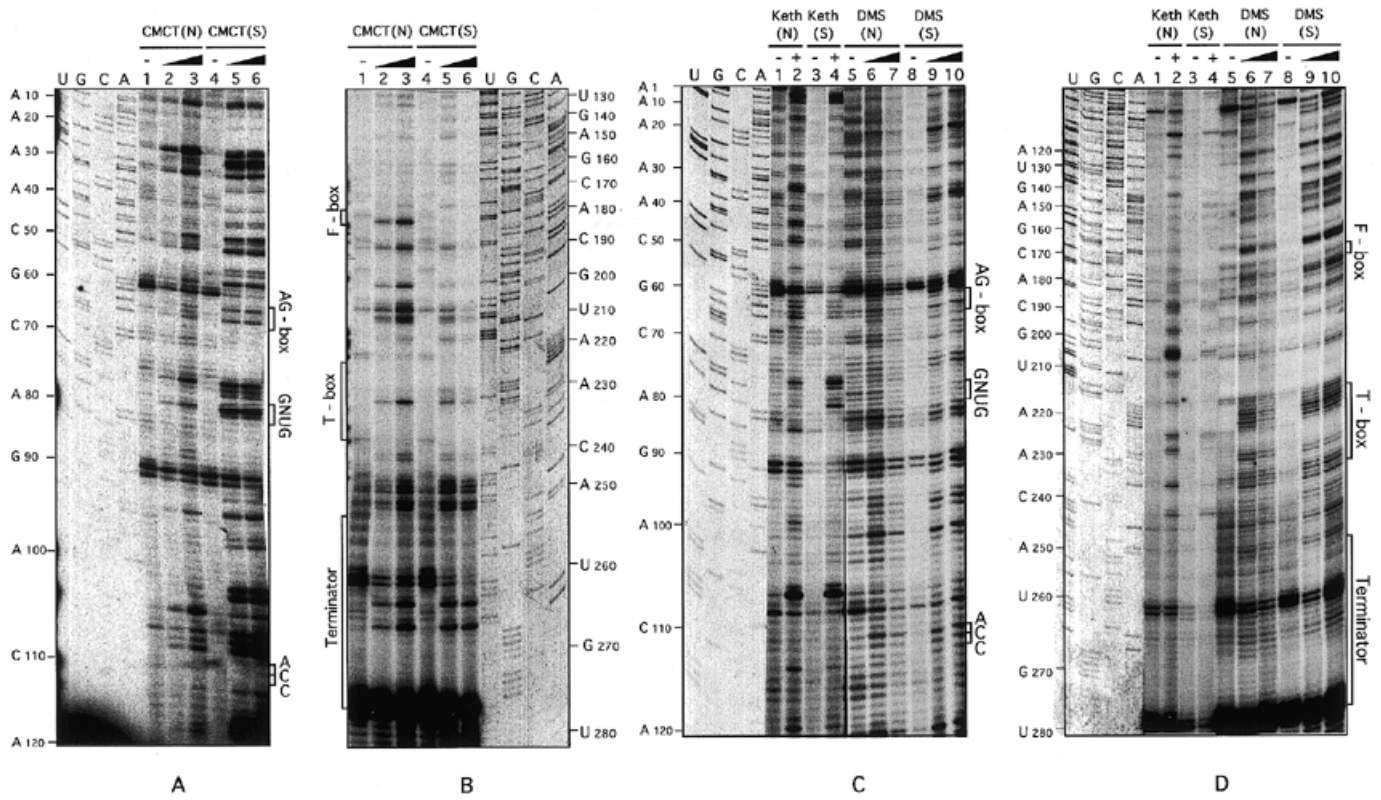

Figure 2. Autoradiograms showing the results of chemical probings of the *thrS* leader region. Transcripts synthesized *in vitro* from plasmid pHMS50 (Materials and Methods) were modified by CMCT, kethoxal (K) and dimethylsulphate (DMS) in the presence (native conditions) or absence (semi-denaturing conditions) of magnesium and analysed by primer extension. The lanes marked U, G, C and A are sequencing reactions using the same primer as in the reverse transcriptase reactions. Lanes marked with plus and minus signs refer to the presence or absence of chemical probes, respectively. Triangles indicate the direction of increasing probe concentration (for DMS) or increasing incubation times (for CMCT) (Materials and Methods). Native and semi-denaturing conditions are referred to as N and S, respectively. The position of conserved sequence or structural elements is indicated to the side of the gels.

RESULTS AND DISCUSSION

In this work, we propose a secondary structure model for the *B.subtilis thrS* leader region. The model is based on the sensitivity of the mRNA to modifications by several chemical probes or nuclease cleavages in solution. We have chosen similar pH, temperature, ionic strength and divalent ion concentrations for the different probes in order to minimize their effects on conformational changes of the RNA.

The RNases T1 and CL3 are single strand-specific enzymes that induce ribose phosphate chain cleavage. RNase T1 specifically cleaves 3′ of guanosine residues and RNase CL3 3′ of adenosine, cytosine and, to a lesser extent, thymidine residues. It is also known as the U-minus enzyme. RNase V1 from cobra venom digests phosphodiester bonds 5′ of residues in double-stranded regions or single-stranded regions which are in a stacking conformation. The use of a combination of these nucleases allows the identification of unpaired, helical or stacked structures.

Probing with DMS, CMCT and kethoxal allows one to test the accessibility of each of the four nucleotides at one of their Watson–Crick positions (19). DMS methylates position N1 of adenine and, to a lesser extent, position N3 of cytosine residues, whereas CMCT modifies position N3 of uridine and, somewhat less, position N1 of guanine residues. Kethoxal reacts with unpaired guanine nucleotides, producing a cyclic adduct between position N1 and its two carbonyl groups. The reactivity of nucleotides towards these chemical probes allows one to distinguish paired and unpaired regions.

Secondary structure of the *thrS* leader region *in vitro*

The *thrS* leader mRNA was transcribed *in vitro* by T7 RNA polymerase using plasmid pHMS50 as template. Its 5′-end corresponds to the last of three possible residues (G) used as transcription start sites by the host polymerase *in vivo* (15). RNAs were incubated with the different chemical and enzymatic probes described above under native (5 mM Mg^{2+}) or semi-denaturing (no Mg^{2+}) conditions. Chemical modification and enzyme cleavages were detected as reverse transcriptase stops in primer extension assays and analysed on polyacrylamide gels. The enzyme cleavages were also detected directly using 3′- and 5′-labelled RNA in order to check for a possible effect of secondary cleavages on the RNA conformation. No such effect was detected (data not shown). The autoradiograms of the chemical modification experiments are shown in Figure 2, those of the RNase probing in Figure 3 and the results of both chemical and enzymatical probings are summarized in Figure 4.

Our results have permitted us to divide the leader sequence into three major domains: the terminator structure, a central domain comprising stem–loops I, II and III, as well as the largely unstructured T-box region, and the specifier domain, preceded by a small 5′-stem–loop structure. The antiterminator structure was

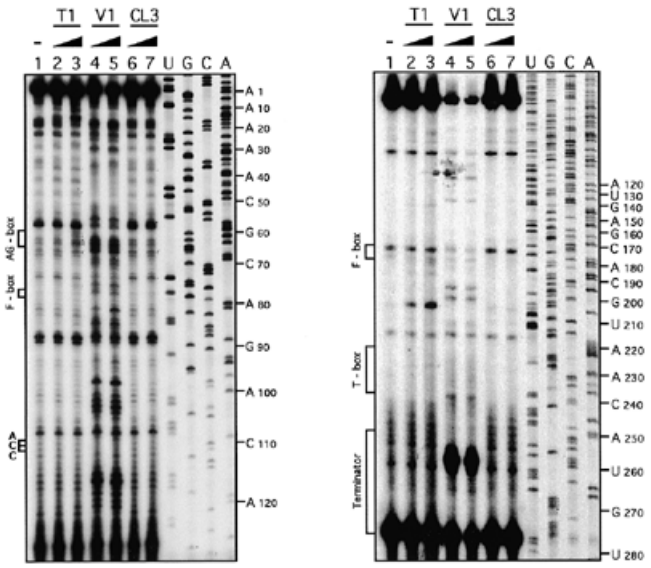

Figure 3. Autoradiograms showing the results of the enzyme probings of the *thrS* leader region. The lanes marked with a minus sign are control reactions in the absence of any RNase. T1, V1 and CL3 represent RNase T1, RNase V1 and RNase CL3 digestions, respectively. Triangles indicate the direction of increasing enzyme concentration. The lanes marked U, G, C and A are sequencing reactions using the same primer as in the reverse transcriptase reactions. The position of conserved sequence or structural elements is indicated to the side of the gels.

probed using a transcript truncated to prevent formation of the terminator.

The terminator. With a calculated free energy of –16 kcal/mol, the terminator stem–loop is the most stable structure (per nucleotide) in the *thrS* leader. None of the nucleotides in the stem were found to be accessible to chemical probes, even under semi-denaturing conditions. A series of RNase V1 cuts on the 5′-side of the terminator helix confirms its stable double-stranded nature (Fig. 4A and B). Strong stop signals are observed at the base of the terminator structure in reverse transcriptase reactions, underlining the stability of the terminator (Fig. 2A and D; Fig. 3). We have already shown the capacity of the terminator to prematurely terminate transcription both *in vivo* and *in vitro* (2,15). Given the stability of the terminator structure, it is clear that the formation of an alternative structure would require additional factors.

Central region. The region between the specifier domain and the terminator forms three stem–loop structures (I, II and III, Fig. 4) and contains a largely unstructured domain centred around the T-box (nt 218–235). The unstructured region between stem–loop III and the terminator comprises 33 nt. Almost all of its residues are accessible to chemical modification under native conditions (Fig. 4A) and there is only one minor RNase V1 cleavage in the entire region. It thus appears that this sequence stretch is extremely flexible, despite a potential complementarity over eight bases, including four G-U pairs. The lack of single strand-specific nuclease cuts suggests, however, that some loose base pairing might occur centred around the two possible G-C base pairs, where the only RNase V1 cleavage was observed (Fig. 4B).

Stem–loop I is a relatively stable structure. None of the principal residues in the stem were accessible to chemical modification under native conditions and RNase V1 cuts were observed at the top and bottom of the stem (Fig. 4). All nucleotides modified under native conditions are either bulged or part of the seven residue loop. The three adenosine residues of the loop are not accessible to cleavage by RNase CL3, which could be due either to steric hindrance or reflect a partial pairing with the two uridine residues of the loop (Fig. 4). While RNase CL3 is generally assumed to be specific for C only, the enzyme preparation used here also cleaved efficiently at A residues (e.g. the apical loops of stem–loop II and the terminator).

Stem–loop II is similar to stem–loop I with a tetraloop and 4 nt bulged out of the lower part of the stem. All single-stranded regions are accessible to chemical modification in the presence of magnesium and are readily cleaved by RNases CL3 and T1 (Fig. 4). Two nucleotides which are based paired in stem–loop II, U_{159} and A_{161}, can be chemically modified under native conditions (Fig. 4A). Their accessibility is probably due to the fact that both residues are at the base of structured regions on either side of the side bulge. The well-conserved F-box sequence ($C_{172}GUUAC_{176}$) is an integral part of stem–loop II. Four of its five residues are involved in base pairing interactions. The experimental data are largely in favour of the structure presented, with the tetraloop and the side bulge being easily accessible to modifications by DMS, kethoxal and CMCT as well as to cleavage by RNases T1 and CL3 (Fig. 4). Point mutations within the F-box sequence almost always have dramatic effects on basal and induced expression of genes of this family (14,20). In the *tyrS* system, where a structure similar to the *thrS* stem–loop II can be formed, compensatory mutations restoring base pairing within the mutated F-box can partially overcome the expression defect (14), underlining the importance of secondary structure in this region. Rollins *et al*. proposed that the possibility of forming an alternative structure, in which the F-box is in an apical loop, was important in the regulation of the *tyrS* gene and suggested that this phenomenon might be generalized to include the other genes of this family (14). We see no evidence that this potential alternative structure forms in the *thrS* leader, either *in vitro* under equilibrium conditions or under inducing conditions *in vivo* (below).

The third structured element in the central region of the *thrS* leader is stem–loop III (Fig. 4). This structure consists of 10 bp, connected by a seven residue loop. Nucleotide U_{196} is bulged. There are two G-U base pairs in the centre of the stem, which contains three reactive residues in the presence of magnesium, G_{194} and G_{195} (kethoxal) and U_{211} (CMCT). While this reflects the weak interaction between G and U residues, this part of the stem appears nevertheless to exist predominantly in a helical conformation as evidenced by the RNase V1 cleavages at the G-U_{211} and the adjacent A-U_{212} base pairs (Fig. 4B).

The specifier domain and the 5′-stem–loop. This is the most complex structure within the *thrS* leader region and is composed of a succession of stems and interior loops. If judged solely by the chemical probing data, the double-stranded regions appear very labile since many guanosine and some adenosine residues in the 5′-half of the structure are accessible to kethoxal and DMS, respectively, under native conditions (Fig. 4A). On the other hand, practically none of the complementary residues in the 3′-half of the structure are modified under the same conditions. There are several reasons, however, why we think the proposed secondary structure is correct. The 3′-half of the specifier domain cannot be folded into a reasonable alternative structure that would

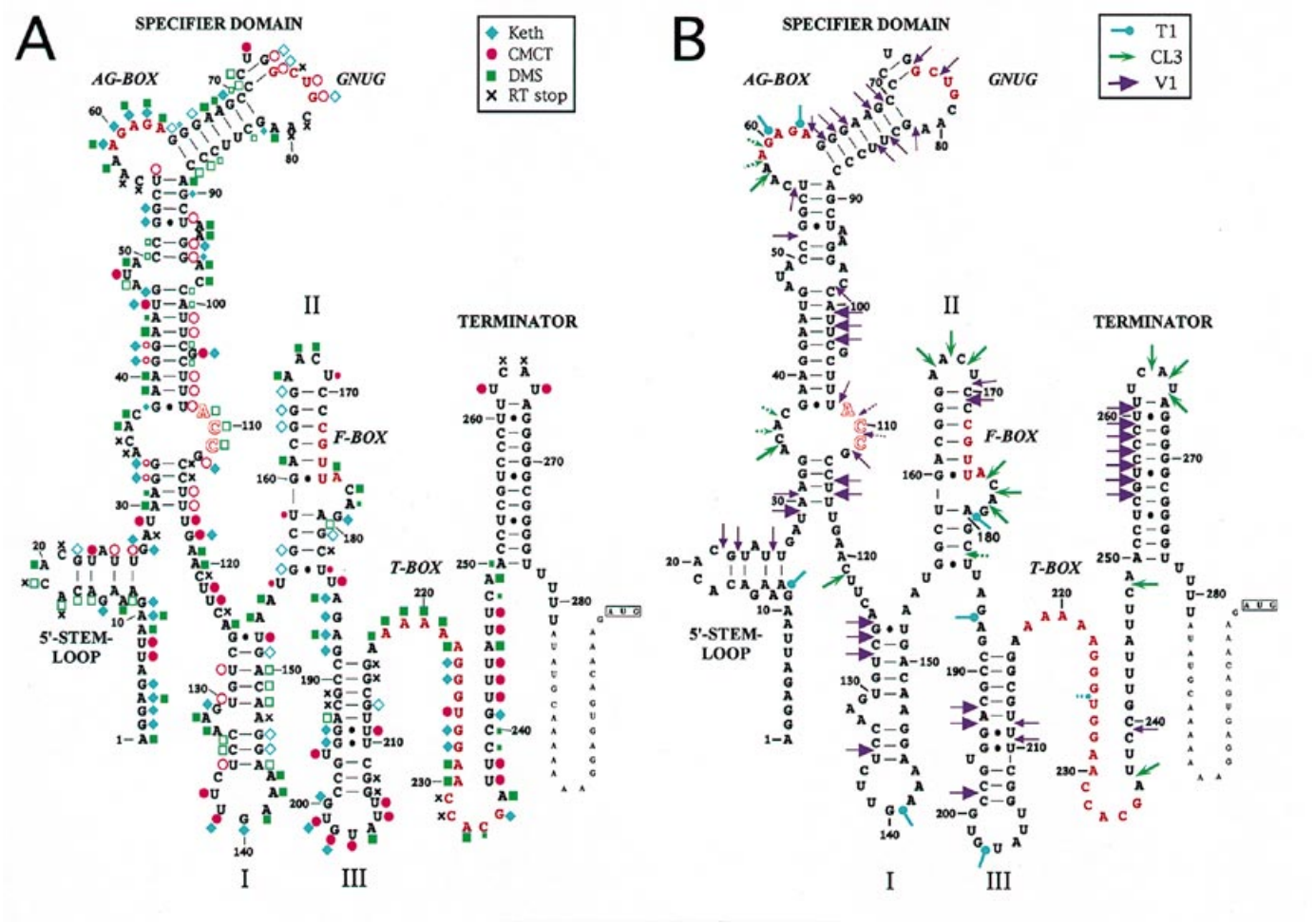

Figure 4. Summary of the results of the chemical (**A**) and enzymatic (**B**) modifications of the *thrS* leader region. The RNA sequence is numbered every 10 bases. The stem–loops are denoted by 5′-stem–loop, specifier domain, stem–loops I, II and III, T-box loop and the terminator. Conserved sequences are in red. The ACC specifier codon is in shadowed letters. Green squares represent DMS modifications, red circles represent CMCT modifications and blue diamonds represent kethoxal modifications. Black crosses indicate non-specific reverse transcriptase stops in the control experiment. The closed and open symbols indicate modifications under native and semi-denaturing conditions, respectively. The arrows in (B) represent the cleavages induced by RNases T1 (blue), CL3 (green) and V1 (purple), respectively. Bases that are not marked are unreactive towards the chemical probes. The size of symbols is proportional to the strength of the modification or cleavage. Broken arrows represent weaker cleavage sites.

leave the 5′-half largely unstructured. Moreover, the nuclease probing experiments support the proposed structure (Fig. 4B), as do the DMS modification data *in vivo*, described below. The RNase V1 cleavage pattern provides evidence in support of a helical conformation of the stems in the upper, central and lower part of the structure (Fig. 4B). We attribute the modification of the G residues, at least partially, to a known denaturing effect of kethoxal under certain conditions (21). The asymmetrical modification pattern would then simply reflect the fact that most of the G residues, by accident or design, are in the 5′-half of the specifier domain.

The specifier domain contains three well-conserved sequence elements that occur in the same relative positions in the leaders of this gene family: the specifier codon (nt A_{109}–C_{111}), the AG-box (A_{59}–A_{63}) and the GNUG sequence G_{75}–G_{78}). As proposed in the model, the specifier codon is indeed bulged, but in contrast to the side loop on the opposite side of the structure, is less readily modified by DMS under native conditions. The degree of its modification can vary slightly from one experiment to the other, as can be judged by comparing Figures 2A and 6. The observation of RNase V1 cleavages within the ACC bulge suggests that these nucleotides exist in a stacking conformation. Such a configuration would increase the stability of the interaction with the similarly structured anticodon of the tRNA, a phenomenon that, for example, explains the surprising stability of tRNA dimers, held together solely by the complementarity of their anticodons (22). The functional significance of the AG-box and the GNUG sequence is still mysterious but the introduction of point mutations has dramatic consequences on both basal and induced expression (C.Condon and H.Putzer, unpublished results; 14). Both consensus sequences appear to be part of loop structures in the upper region of the specifier domain. The AG-box is readily modified by DMS and kethoxal under native conditions and is also easily accessible to RNases T1 and CL3. On the other hand, the apical loop containing the GNUG sequence is much more structured, is modified mainly under semi-denaturing conditions, is not accessible to single strand-specific nucleases and is

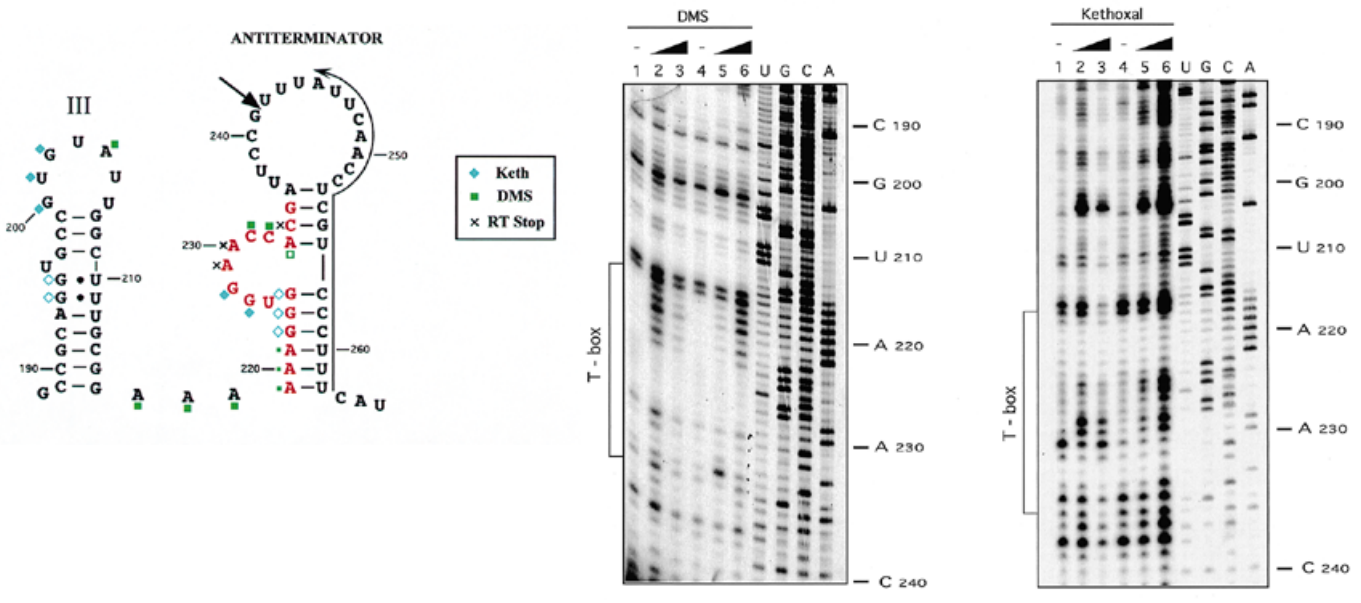

Figure 5. Autoradiogram and summary of the results of the chemical probing of the antiterminator in the *thrS* leader region. The right hand side shows the autoradiograms of the chemical modifications. The modified bases were revealed by reverse transcription using oligonucleotide HP106 as primer (Materials and Methods). The sequence complementary to the oligonucleotide is sidelined. The left hand side summarizes the results of the chemical modifications. Green squares represent DMS and blue diamonds kethoxal modifications. Black crosses indicate non-specific reverse transcriptase stops in the control experiment. The T-box is in red. The arrow indicates the processing site seen *in vivo*.

susceptible to cleavage by RNase V1 (Fig. 4). The reactivity pattern of the GNUG loop thus resembles that of the specifier codon bulge and may be indicative of a stacked conformation or a long-range tertiary interaction. A potential target sequence suitable for this kind of tertiary interaction can be found between the specifier domain and stem–loop I ($C_{124}AGC_{127}$). However, compensatory mutations introduced into this sequence in order to restore the non-inducible phenotype of point mutations in the GCUG sequence showed no effect at all (D.Luo and H.Putzer, unpublished data). Moreover, Pb(II) probing of the *thrS* leader did not reveal any significant tertiary interactions (data not shown). We thus think that the modification pattern of the GNUG loop observed *in vitro* is rather due to a stacking conformation.

We have identified a small secondary structure, termed the 5′-stem–loop, between the transcription start site and the base of the specifier domain (Fig. 4). At first glimpse this stem–loop does not appear to be very stable or of much consequence, but the chemical modification and RNase V1 cleavage patterns clearly support its existence. We cannot propose a functional role for this structure, but we can identify similar potential stem–loops 5′ of the specifier domain in transcripts of practically all genes of this family in *B.subtilis*. They vary considerably in their size and stability but always contain a high percentage of A + U residues in both the double-stranded and looped regions. This is especially apparent in the large potential 5′-stem–loop of the *B.subtilis trpS* leader, where 16 out of 20 bp in the stem are A-U base pairs (data not shown).

The antiterminator. Genetic evidence obtained with three different genes of the T-box family (*tyrS*, *thrS* and *ilv-leu*) strongly indicates that, during amino acid starvation, the -NCCA-3′ end of the uncharged tRNA interacts with the complementary 5′-UGGN′-3′ sequence of the T-box bulged out of the proposed antiterminator structure (Fig. 1; 3,4,10). This interaction is thought to favour formation of the antiterminator, which otherwise probably does not exist given the much greater stability of the mutually exclusive terminator structure. In the case of the *thrS* leader, where the difference in ΔG between the two structures is 11 kcal, this is clearly confirmed by the probing experiments (Fig. 4). In order to study the conformation of a potential antiterminator structure, we thus used T7 RNA polymerase to synthesize a transcript *in vitro* that ends at the *Eco*NI site within the loop of the terminator. Such a shortened transcript contains all the sequences necessary to form the proposed antiterminator structure, but not the competing terminator. Moreover, this truncation does not alter the reactivity pattern of any of the nucleotides upstream of the potential antiterminator (data not shown). The chemical modification profile between residues C_{190} and C_{240} of the truncated *thrS* leader is in agreement with the current regulatory model (Fig. 5). The guanosine (G_{227} and G_{228}) and cytosine (C_{231} and C_{232}) residues of the antiterminator bulge are modified by kethoxal and DMS, respectively, under native conditions. On the other hand, the three guanosine residues G_{223}–G_{225} and A_{233} making up the central part of the stem are only accessible to modification in the absence of magnesium, i.e. under semi-denaturing conditions. Since the oligonucleotide used to prime the reverse transcriptase reactions hybridizes to the 3′-end of the transcript, we could not analyse the entire antiterminator structure. Due to the length of the transcript and the inconsistent migration of very short RNA fragments we could not reliably extend the data collection by direct (3′ or 5′) labelling methods. Nevertheless, the structure proposed by Grundy and Henkin (1), and shown in Figure 5, is favoured over an alternative stem–loop structure that can theoretically form in the shortened *thrS* leader. In this, theoretically more stable, structure a large bulge forms in the 3′-half of the stem

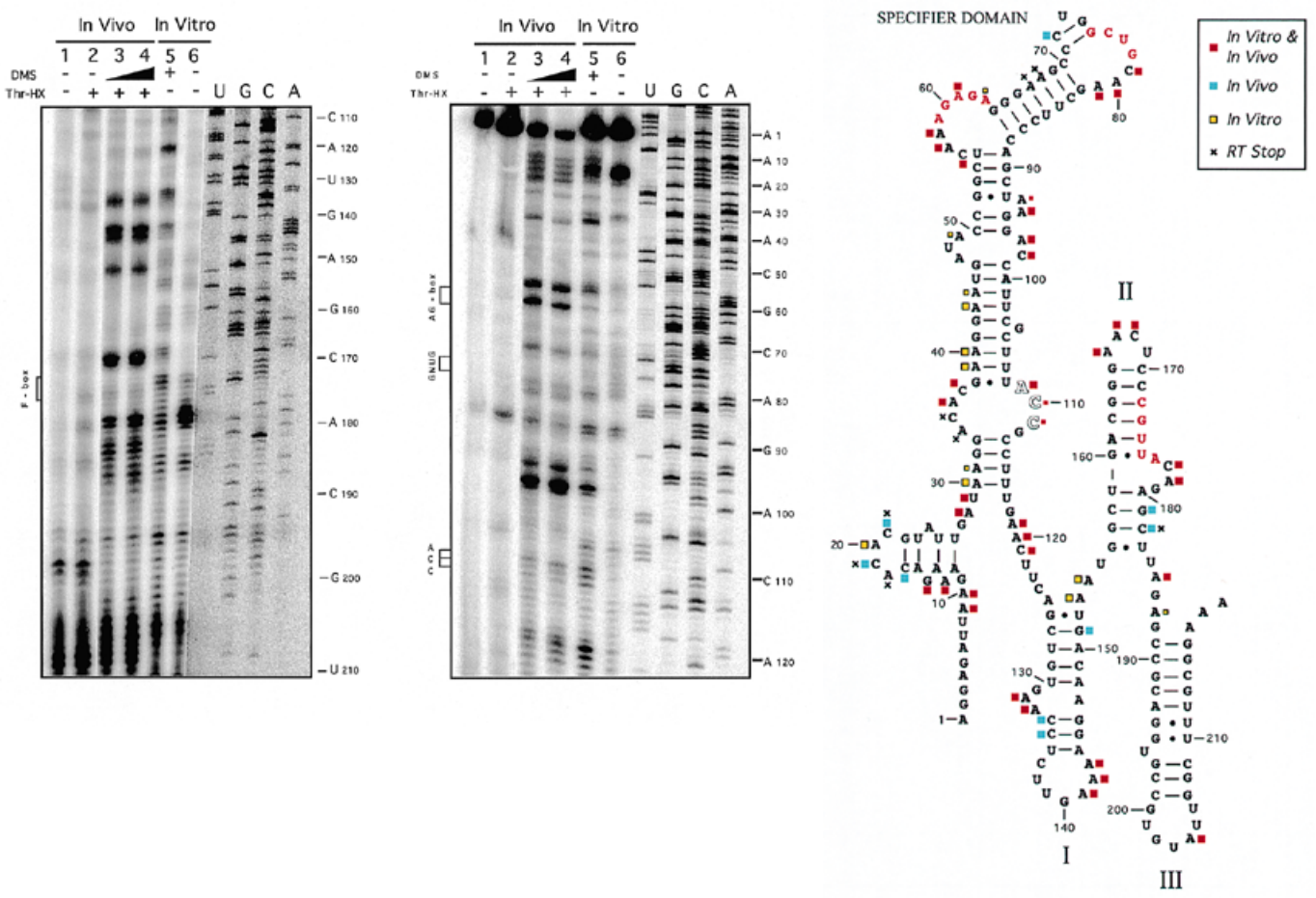

Figure 6. Comparison of the *in vivo* and the *in vitro* DMS modifications of the *thrS* leader region. Thr-HX stands for threonine hydroxamate which was added to the culture to starve it of threonine. For the *in vivo* tests, bulk RNA was isolated from strain HP48 under normal growth conditions (without threonine hydroxamate) or under threonine starvation conditions (in the presence of threonine hydroxamate). Control reactions were done in the absence of DMS under both normal (lane 1) and threonine starvation conditions (lane 2). Two concentrations of DMS (100 and 200 µl/20 ml culture) were used for the tests *in vivo*. The direction of increased concentration is represented by the triangle. For the *in vitro* tests, only one concentration of DMS was used. The presence or absence of DMS is denoted by a + or – sign. Lanes marked U, G, C and A are sequence reactions using the same oligonucleotide as primer as were used in the reverse transcriptase reactions. The consensus sequences in the leader region of the T-box genes are indicated in red. The results of the comparison are summarized in the schematic secondary structure of the *thrS* leader. The RNA sequence is numbered every 10 bases. The yellow, blue and red squares indicate the DMS modifications observed *in vitro* only, *in vivo* only or both *in vivo* and *in vitro*, respectively. A black cross marks a non-specific reverse transcriptase stop in the control experiment. The terminator stem–loop and the T-box region were not probed in the *in vivo* experiments, because of the processing of the *thrS* messenger RNA that occurs in this region. The processing site is located at nt 241. The primer HP241 used is complementary to sequences just upstream of this site.

and the T-box sequence is in a completely double-stranded conformation.

DMS modification of the *thrS* leader *in vivo*

Structural probing *in vitro* necessarily involves an RNA molecule whose conformation represents a state of minimal or near minimal free energy. In order to assess the significance of the data obtained *in vitro* described above, we probed the *thrS* leader structure *in vivo* using DMS. While the comparison is limited to only one probe, DMS being the only one of the probes used capable of penetrating the bacterial membrane, this method nevertheless provides valuable information on RNA structure during the transcription process itself and in the presence of any potential regulatory factors absent *in vitro*. We could only analyse the *thrS* leader up to nt 210 because of an endonucleolytic cleavage event that occurs at position 241 *in vivo*. Almost 90% of all *thrS* transcripts are cleaved at this position when expression is induced under threonine starvation conditions (11,12). We thus chose a reverse transcriptase primer which hybridizes to the 3′-end of the processed transcript. The DMS modifications observed under threonine starvation conditions *in vivo* essentially confirm the *in vitro* data (Fig. 6), although the specifier domain appears to be significantly more stable *in vivo*. None of the adenosine residues in the double-stranded regions, which were quite accessible to methylation by DMS *in vitro*, were modified *in vivo* (Fig. 6). At the same time the majority of bulged and looped regions showed a very similar modification pattern *in vivo* and *in vitro*. While the ACC specifier codon was not always modified *in vitro*, we reproducibly obtained significant modification *in vivo*. The central domain containing stem–loops I–III also showed a very similar reactivity pattern towards DMS *in vivo* and *in vitro*, as can be judged by the accessibility of single-stranded regions and the non-reactivity of helices (Fig. 6).

Residues in single-stranded regions which reacted with DMS *in vitro* under native conditions but were not modified *in vivo* are of special interest, since they might indicate sites of interaction with potential regulatory factors. However, there are relatively few of such nucleotides: A_{20}, A_{49}, A_{59}, A_{63}, A_{153} and A_{154} (Fig. 6). Thus, if an interaction with regulatory factors exists, it involves only very few nuleotides or the time of interaction is short relative to the half-life of the mRNA. Given that we don't see any evidence for the tRNA:mRNA interaction either, even under starvation conditions *in vivo*, we favour the hypothesis that these interactions may be short lived.

It is interesting to note that some residues in supposedly double-stranded regions, which are not reactive to DMS under native conditions *in vitro*, are modified *in vivo*. These are C_{17}, C_{72}, C_{134}, C_{135}, G_{151} and G_{181}, even though guanosine is normally not modified by DMS *in vitro* (Fig. 6). These differences might reflect the formation of intermediary structures in the leader as it is being synthesized *in vivo* or the binding of other factors. The non-reactivity of A_{153} and A_{154}, and the accessibility of G_{181} to DMS modification *in vivo* could, for example, be explained by alternative folding of stem–loop II. The stem can be shifted by 3 nt in the 5′ direction to include nt A_{153}–U_{155}. This structure would contain a five residue bulge (nt 158–162), be compatible with the observed modifications and render G_{181} accessible. In both cases, the F-box element is involved in base pairing interactions. DMS modification of guanosine residues leading to chain termination occured only *in vivo* and has been observed previously (18). This may reflect the formation of a ring-opened derivative of 7-methylguanine found in cells treated with alkylating agents (23) which was shown to produce premature termination in primer extension reactions using T4 DNA polymerase (24).

We also compared folding of the *thrS* leader *in vivo* under normal and inducing conditions. This study was limited to the specifier domain, since we did not obtain clear signals for the central domain under normal growth conditions. During threonine starvation, significant increases in DMS reactivity occur at numerous residues throughout the specifier domain (Fig. 7). In the region comprising the first 85 nt or so, it is difficult to assess their significance because of an apparent disappearance of the background signal from this point on in lanes 4 and 5. There are two possible explanations for this phenomenon. The first is that the region between +1 and +85 is more sensitive to nucleolytic attack under normal growth conditions and that the decreased stability accounts for the lack of signal. An increased sensitivity of this part of the leader to nucleases under non-inducing conditions is consistent with the observation that the signal corresponding to the transcription start point is 3- to 4-fold stronger under starvation conditions relative to non-inducing conditions in the control experiment (without DMS), despite the fact that we have previously shown that regulation of *thrS* expression does not occur at the level of transcription initiation (2,15). A second possibility (that does not account for the increased transcription start point signal) is that these modifications are significant and the increased accessibility of these residues during induction is due to the dissociation of potential regulatory factors that bind these regions under normal growth condtions. Experiments are underway to distinguish between these two possibilities.

Downstream of nt 85 the background signal is more or less the same under all growth conditions and several residues are nonetheless significantly more reactive to DMS during threonine

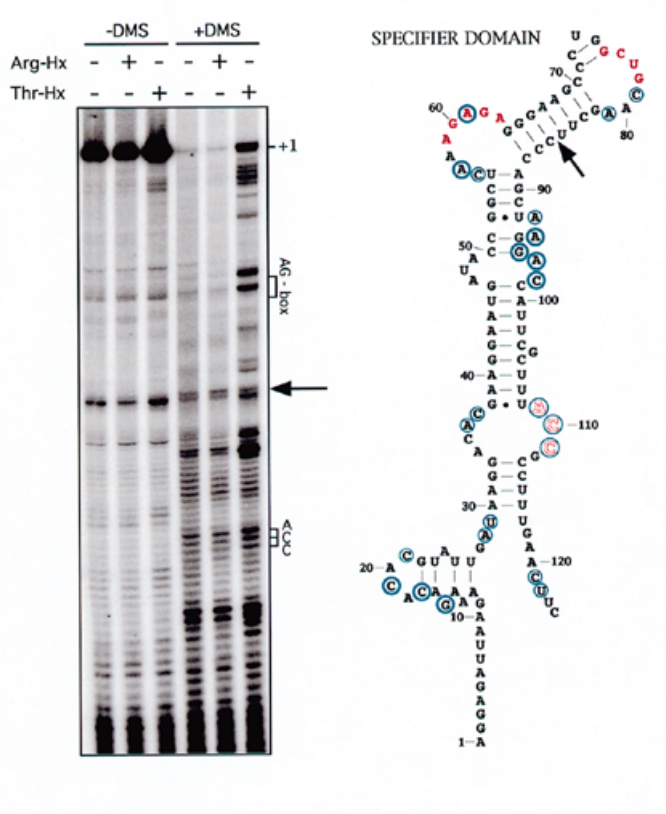

Figure 7. Comparison of the DMS modification pattern *in vivo* under normal growth and threonine or arginine starvation conditions. The autoradiogram shows the DMS modifications revealed by reverse transcription using primer HP9 complementary to nt 145–165. The presence or absence of threonine hydroxamate (Thr-HX) and arginine hydroxamate (Arg-HX) in the cell cultures is indicated by + or – signs. Bases corresponding to the specifier codon, the AG-box and the transcription start point are indicated to the right side of the autoradiogram. The arrow indicates the approximate position in the leader up to which the 5′-half of the specifier domain shows a differential background modification pattern under different conditions (text). Blue circles highlight nucleotides which are specifically more sensitive to DMS modification under threonine starvation conditions.

starvation (Fig. 7). This observation, which also includes the ACC specifier codon, could be related to the fact that the entire domain seems to be much better defined *in vivo* than *in vitro*. The increased compactness of the structure, possibly a result of interaction of the mRNA with the tRNA and/or other regulatory factors, might force the bulged regions into a more reactive conformation. On the other hand, it is also possible that the increased modification of these bands is caused by the dissociation of potential protein factors under starvation conditions, as described above.

Conclusion

Our data provide insight into the structural elements of the *thrS* leader mRNA involved in tRNA-mediated antitermination. Many, if not all, of the leaders of this gene family can be folded in the same or a very similar manner and this holds especially true for the specifier domain (data not shown). Despite several attempts, we could not reproducibly demonstrate a direct interaction between the *thrS* leader and $tRNA^{Thr}_{GGU}$ isoacceptor. This was not entirely unexpected since uncharged $tRNA^{Thr}_{GGU}$

alone is insufficient to promote increased termination read-through in an *in vitro* transcription system. One of the reasons might be the lack of stability of the specifier domain *in vitro* or the absence of significant tertiary structure as judged by lead (PbII) probing (data not shown). The interaction with the tRNA alone is apparently not sufficient to fold the leader mRNA into a higher order structure which might well occur *in vivo*. The identification of other factors involved in this complex regulatory mechanism is needed to attribute a more precise function to the different structural elements.

ACKNOWLEDGEMENTS

We thank Pascale Romby and Eric Westhof for helpful discussions. This work was supported by the Centre National pour la Recherche Scientifique (UPR9073); D.L. was the recipient of a BDI scholarship from the CNRS.

REFERENCES

1 Grundy,F.J. and Henkin,T.M. (1993) *Cell*, **74**, 475–482.
2 Putzer,H., Gendron,N. and Grunberg-Manago,M. (1992) *EMBO J.*, **11**, 3117–3127.
3 Putzer,H., Laalami,S., Brakhage,A.A., Condon,C. and Grunberg-Manago,M. (1995) *Mol. Microbiol.*, **16**, 709–718.
4 Marta,P.T., Ladner,R.D. and Grandoni,J.A. (1996) *J. Bacteriol.*, **178**, 2150–2153.
5 Grundy,F.J., Haldeman,M.T., Hornblow,G.M., Ward,J.M., Chalker,A.F. and Henkin,T.M. (1997) *J. Bacteriol.*, **179**, 3767–3772.
6 Raya,R., Bardowski,J., Andersen,P.S., Ehrlich,S.D. and Chopin,A. (1998) *J. Bacteriol.*, **180**, 3174–3180.
7 Grundy,F.J. and Henkin,T.M. (1994) *J. Mol. Biol.*, **235**, 798–804.
8 Henkin,T.M. (1994) *Mol. Microbiol.*, **13**, 381–387.
9 Putzer,H., Grunberg-Manago,M. and Springer,M. (1995) In Söll,D. and RajBhandary (eds), *tRNA: Structure, Biosynthesis and Function.* ASM Press, Washington, DC, pp. 293–333.
10 Grundy,F.J., Rollins,S.M. and Henkin,T.M. (1994) *J. Bacteriol.*, **176**, 4518–4526.
11 Condon,C., Putzer,H. and Grunberg-Manago,M. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 6992–6997.
12 Condon,C., Putzer,H., Luo,D. and Grunberg-Manago,M. (1997) *J. Mol. Biol.*, **268**, 235–242.
13 Kunst,F. *et al.* (1997) *Nature*, **390**, 249–56.
14 Rollins,S.M., Grundy,F.J. and Henkin,T.M. (1997) *Mol. Microbiol.*, **25**, 411–21.
15 Gendron,N., Putzer,H. and Grunberg-Manago,M. (1994) *J. Bacteriol.*, **176**, 486–494.
16 Harwood,C.R. and Cutting,S.M. (eds) (1990) *Molecular Biological Methods for Bacillus.* John Wiley & Sons Ltd, Chichester, UK.
17 Luo,D., Leautey,J., Grunberg-Manago,M. and Putzer,H. (1997) *J. Bacteriol.*, **179**, 2472–2478.
18 Mayford,M. and Weisblum,B. (1989) *EMBO J.*, **8**, 4307–4314.
19 Ehresmann,C., Baudin,F., Mougel,M., Romby,P., Ebel,J.P. and Ehresmann,B. (1987) *Nucleic Acids Res.*, **15**, 9109–9128.
20 Grandoni,J.A., Fullmer,S.B., Brizzio,V., Zahler,S.A. and Calvo,J.M. (1993) *J. Bacteriol.*, **175**, 7581–7593.
21 Jaeger,L., Westhof,E. and Michel,F. (1993) *J. Mol. Biol.*, **234**, 331–346.
22 Houssier,C. and Grosjean,H. (1985) *J. Biomol. Struct. Dyn.*, **3**, 387–408.
23 Beranek,D.T., Weis,C.C., Evans,F.E., Chetsanga,C.J. and Kadlubar,F.F. (1983) *Biochem. Biophys. Res. Commun.*, **110**, 625–631.
24 O'Connor,T.R., Boiteux,S. and Laval,J. (1988) *Nucleic Acids Res.*, **16**, 5879–5894.