Nucleic Acids Research, Vol. 19, No. 13 3653–3660

Conformation and structural fluctuations of a 218 nucleotides long rRNA fragment: 4-thiouridine as an intrinsic photolabelling probe

Yolande Lemaigre Dubreuil, Alain Expert-Bezançon[1] and Alain Favre
Institut Jacques Monod, CNRS-Université Paris 7, Tour 43, 2 place Jussieu, 75251 Paris Cedex 05 and [1]Centre de Génétique Moléculaire, CNRS-Université Pierre et Marie Curie, 91190 Gif sur Yvette, France

Received March 12, 1991; Revised and Accepted May 28, 1991

ABSTRACT

The structure in solution of an RNA fragment (218 nucleotides long) containing part of *E. coli* 16S rRNA domain 2 has been studied using the intrinsic photoaffinity probe 4-thiouridine (s^4U). *In vitro* transcription with T7 polymerase, in the presence of s^4U triphosphate yielded complete RNA molecules. An affinity electrophoresis system based on Phenylmercuric substituted polyacrylamide (APM) gels allows separation of the RNA chains as a function of their s^4U content. Distribution of s^4U within chains follows a binomial law indicating that (i) substitution is close to random, (ii) efficiency of s^4U incorporation is 0.22 times that of U. The monothiolated RNA fraction isolated from APM gel was irradiated at 366 nm under native conditions and the intramolecularly crosslinked molecules, (34%), were separated on denaturing polyacrylamide gel according to loop size. The positions of the two partners of bridges were identified by mean of reverse transcription and RNA sequencing. 17 of the 41 possible s^4U positions lead to detectable bridges. These crosslinks formed efficiently at the border of bihelical regions or when structural mobility is allowed. The pattern of crosslinks is in agreement with the previously proposed secondary structure but indicates that it is much more flexible than expected.

INTRODUCTION

Several experimental methods are available for the analysis of secondary and tertiary structures of long RNA molecules. Chemical and enzymatic probing generally give information on secondary structure and protein-RNA interactions. Directed mutagenesis and chemical and photochemical crosslinking can yield information on long range tertiary structures and alternative structures that are not accessible by other means. UV (254 nm), chemical (1, 2) and psoralen crosslinking have already produced valuable information on tertiary folding of 16S rRNA and mRNAs (3).

Potential advantages of s^4U discussed in (4) are: i) it is able to substitute U in RNA chains both *in vitro* and *in vivo*, ii) it can be selectively photoactivated at 330–366 nm, iii) it is able to form photocrosslinks with pyrimidines but also with purines, the reactivity order for collisional photocoupling being U>A>C>G (5), iv) it efficiently yields long range crosslinks both in cis e.g. the 8–13 link in tRNA and in trans e.g. between mRNA analogues and 16S rRNA (4, 6, 7).

The model RNA we have chosen, a 181 nucleotide segment of domain 2 of *E. coli* 16S rRNA, is one of the most studied RNA regions involved in multiple interactions. Its structure in naked 16S rRNA, in the ribosome or as a fragment with and without proteins S8 or S15 has been studied by different methods: chemical and enzymatic probing (8, 9, 10, 11), directed mutagenesis (12). It is mainly composed of two helices, **H** and **H'** (Fig. 1), and a three dimensional structure was proposed based on chemical and enzymatic analysis (10). It is the site of interaction with proteins S8 and S15 and also contains sites of interaction or contact with other proteins S11, S21 (helix **H'**), S17 (helix **H**) (1), release factor 2 (RF2) (13), tRNA (14) and mRNA (7).

An analysis of this model RNA was started using s^4U monosubstituted molecules which, after crosslinking, were separated by electrophoresis on a denaturing gel according to loop size. The 3′ sides of loops were then identified by reverse transcription and the 5′ sides either by a second primer extension (for long loops) or by alkaline hydrolysis of 5′ end labelled molecules (for short loops). We do not detect interactions between the two helices but instead confirm the secondary structure. In addition this study shows that it is possible to detect structural fluctuations of RNA molecules in solution and allows us to define some characteristics of s^4U photochemistry in native RNA.

MATERIALS AND METHODS

Chemicals and enzymes

[γ^{32}P]ATP (185TBq/mmole) was from Amersham. Acrylamide and N,N′-methylene-bisacrylamide were from BDH. Nucleotides, deoxynucleotides and dideoxynucleotides were from Pharmacia. s^4UTP was synthesized according to Scheit (15) with modifications (gift of J.L. Fourrey). Qiagen columns were from Diagen. N-acryloylaminophenylmercuric chloride (APM)

was a generous gift of G.L.Igloï. Deoxyoligonucleotides were synthesized with an Applied Biosystem apparatus using the phosphoramidite method.

Restriction enzyme, phage T7 RNA Polymerase and Avian Myeloblastosis Reverse Transcriptase were from Boehringer. Bovine Serum Albumin was from BRL and Placental Ribonuclease Inhibitor (RNasin) from Promega Biotech. RNases T1 and U2 and RNA Ligase were from Pharmacia. Calf Intestine Alkaline Phosphatase was from Stratagene and phage T4 Polynucleotide Kinase from Biolabs.

Preparation of the RNA fragment

In vitro transcription with s^4UTP. The DNA fragment corresponding to the 16S rRNA segment (nucleotides 578 to 758 in the 16S rRNA sequence) was inserted in plasmid pT7.2 (10) and linearized by Bam H1. A low specific activity was chosen to limit the photolysis of s^4U by Cerenkov radiation (16) during the preparation of samples before photoactivation. After one hour of transcription at 37°C, the RNA was purified from DNA matrix, proteins and free nucleotides on an affinity column (Qiagen) according to the instructions of the manufacturer. The synthesized 16S rRNA fragment contains additional sequences at both extremities, resulting from transcription of part of the polylinker. They correspond to the 28 nucleotides located between the transcription start and the 5′ end of the insert (pppGGGAG-ACCGGAAGCUUGGGCUGCAGGUC) and to the 9 nucleotides between the 3′ end of the insert and the Bam H1 cut (AGACGGAUC).

Affinity polyacrylamide gel electrophoresis. Thiolated RNA molecules were electrophoresed on polyacrylamide gels which contained N-acryloylaminophenyl mercuric chloride (APM) (16).The transcribed RNA was diluted with a volume of loading buffer (deionised formamide, 0.025% (w/v) xylene cyanol) heated for 1 min at 65°C and loaded on a 12% (w/v) polyacrylamide slab gel containing 7 M Urea, 2 μg/ml APM which was prepared in the electrophoresis buffer: 0.089 M Tris-borate, pH 8.3, 2.5 mM Na_2Edta.

Photocrosslinking of the RNA fragment and separation of photoproducts

Photocrosslinking of monothiolated molecules. Irradiation of the samples was performed with the 366 nm line ($\Delta\lambda$ = 10 nm) of an HBO 200 W superpressure mercury lamp selected by mean of a Baush and Lomb monochromator. A Schott WG 345 filter was added at the exit slit of the monochromator and the optical cuvette was placed 1 cm from it. Purified monothiolated RNA (0.6 pmol/μl in 20 mM Tris-acetate, pH 7.6, 300 mM Potassium acetate, 20 mM Magnesium acetate buffer) was first incubated for 15 min at 42°C, then rapidly cooled and irradiated for 30 min at 4°C.

Preparative denaturing polyacrylamide gel electrophoresis. Irradiation products were precipitated with ethanol, resuspended in loading buffer (deionised formamide, 0.025% (w/v) xylene cyanol) and analysed by electrophoresis on a 12% (w/v) polyacrylamide slab gel (acrylamide to bis(acrylamide) weight ratio 29:1) containing 7 M Urea, prepared in 0.089 M Tris-borate, pH 8.3, 2.5 mM Na_2Edta. Prior to loading, the samples were heated for 1 min at 65°C. Electrophoresis was conducted at 60 volts/cm for 3 hours at 50°C.

Reverse transcription analysis

Primer extension. Three primers, noted 1, 2 and 3, were used. Their sites of complementarity are: 1 (C754-3′OH extremity), 2 (A694-U707) and 3 (A622-U636). They were labelled with [γ^{32}P]ATP at their 5′ end according to the method of Silberklang et al. (17) and hybridized with modified or unmodified RNA. The extension reaction was conducted according to established procedures (18).

Purification of RNA–DNA hybrids by non denaturing gel electrophoresis. After transcription of heterogeneous crosslinked RNA molecules with primer 1 and reverse transcriptase, elongation mixtures were diluted with one volume of 8 M Urea-0.025% (w/v) xylene cyanol and loaded on 12% gel slabs (400×200×0.4mm, acrylamide to bis(acrylamide) weight ratio 29:1) prepared in the electrophoresis buffer: 0.089 M Tris-borate, pH 8.3, 2.5 mM Na_2Edta. Electrophoresis was carried out at 10°C for 5 hours (short run separation) or overnight at 37.5 V/cm. After autoradiography, the samples were eluted and analysed on sequencing gels, either directly or after elongation with a second primer.

Limited alkaline hydrolysis analysis

5′ end labelling. Prior to labelling with [γ^{32}P]ATP at their 5′ end according to (17), RNA molecules were dephosphorylated with Calf Intestine Alkaline Phosphatase (CIAP) (18). The labelled RNA was purified by electrophoresis on (400×200×0.4mm) 10% (w/v) polyacrylamide slab gels

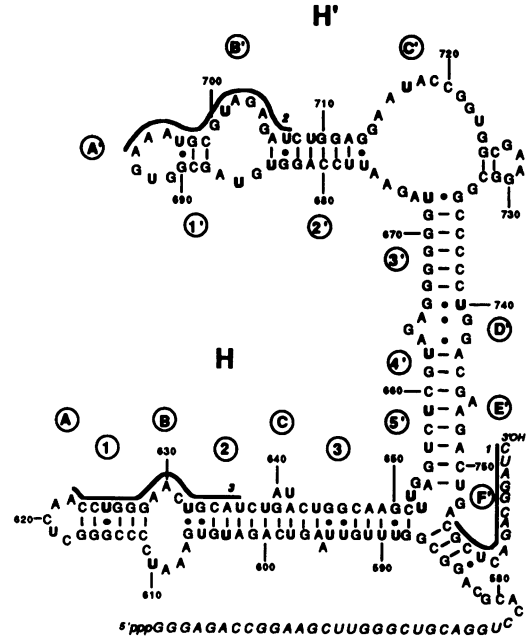

Figure 1. Secondary structure of the 218 nucleotides long RNA transcript used in this study. The 28 nucleotides at the 5′ end and the 9 nucleotides at the 3′ end (*italics*) are transcribed from the polylinker of PT7.2, the section of 16S rRNA domain 2 (181 nucleotides) extends from C578 to C758. The 41 Us are indicated in bold faced lettering and those found crosslinked in outlined lettering. The three DNA primers used for reverse transcription are indicated by bold line. For convenience, in helices **H** (G588 to C651) and **H′** (A655 to U751), secondary motives are numbered: **A**, end loop; **B** and **C**, internal loops of **H**. **A′**, end loop; **B′**, **C′**, **D′**, **E′**, **F′** internal loops or bulges of **H′**.

containing 7 M Urea, prepared in the electrophoresis buffer and run at 40 W and 50°C for 2 hours.

Alkaline hydrolysis and base specific cleavage. To determine the exact nucleotide involved in a bridge, limited alkaline hydrolysis and enzymatic sequencing were done according to J.M. D'Alessio (19).

Sequencing gel. (520×200×0.4mm) 7% (w/v) polyacrylamide slab gels (acrylamide to bis(acrylamide) weight ratio 19:1) containing 7 M Urea were prepared in the electrophoresis buffer: 0.089 M Tris-borate, pH 8.3, 2.5 mM Na_2Edta. During preelectrophoresis (45 min, 60 W, 65°C), a saline gradient was established in the gels by addition in the lower buffer of 1 M Sodium acetate. The gel was run at the same power and temperature for 2 or 2.5 hours.

RESULTS

s^4U incorporation by T7 RNA polymerase

Transcription with T7 polymerase was performed in the presence of a constant starting concentration of NTPs (fixed at 1.25 mM) including UTP+s^4UTP. The efficiency of transcription, i.e. the number of RNA copies per DNA template decreased from a mean value of 250 to 120 when the ratio R=s^4UTP/s^4UTP+UTP increased from 0 to 1. Hence T7 polymerase is able to synthetise fully thiolated molecules and its activity is not impaired by the presence of successive Us in contrast to *E. coli* polymerase (20). After removal of DNA and proteins on a Qiagen column, the synthesized RNA fragments migrate as a single band on a denaturing gel. However they can be fractionated as a function of their s^4U content by affinity electrophoresis on APM gel. The thiol containing nucleic acids are retarded by specific interaction with an organomercurial derivative (APM) included in the gel (16). As shown in Fig.2a the s^4U substituted fragments yield discrete bands numbered 0, 1, 2... Retardation increases with R and fully substituted RNA (lane h) does not enter the gel. Obviously band 0 corresponds to non substituted RNA. Bands were identified using a large scale RNA synthesis (5 A_{260} units). With R = 0.1, the mean substitution level is 0.95 ± 0.05 s^4U per chain as determined from the absorbance ratio at 330 and 260 nm (21). After APM gel fractionation, the s^4U contents of bands 0 and 1 were determined by the same procedure and found to be respectively 0 and 1 ± 0.1. This determination was extended to higher values (≤ 4) of i, the number of s^4U residues incorporated per chain, as follows. The normalized weight W_i of band i, $\Sigma W_i = 1$, obtained by Cerenkov counting is such that $0.95 = \Sigma\, i\, W_i$ ($W_0 = 0.36$, $W_1 = 0.415$, $W_2 = 0.18$, $W_3 = 0.05$). The data allow evaluation of the efficiency of s^4U incorporation relative to U, i.e. $(1-R)\, \Sigma\, i\, W_i / R\, (41 - \Sigma\, i\, W_i)$, which is found to remain constant (0.22 ± 0.2) when R varies between 0.05 and 0.25. Finally the experimental distribution of chains containing 0, 1, ...i s^4U fits closely the theoretical distribution given by a binomial law $W'_i = C^i_{41}\, q^i\, (1-q)^{41-i}$ indicating that, for $i \leq 4$, incorporation is close to random (Fig.2b).

We have observed the concentration dependent formation of dimers immediately after T7 transcription, independently of the presence of s^4U in the molecules. Monomers and dimers can be separated on sucrose gradients or on non denaturing polyacrylamide gels (data not shown). The low concentration used in the renaturation step limits dimer formation to a minimum (10%) but does not eliminate it completely.

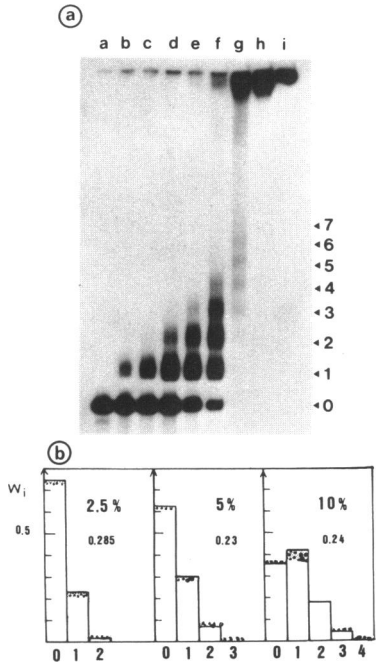

Figure 2. Separation and distribution of thiolated RNA molecules. **a.** RNA molecules synthesized *in vitro* by T7 RNA polymerase with increasing concentrations of s^4UTP, at constant (s^4UTP+UTP) = 1.25 mM, are separated on APM gel according to the number of s^4U residues per molecule. Lanes a to i, R=0, (a); 0.01, (b); 0.025, (c); 0.05, (d); 0.1, (e); 0.2, (f); 0.5, (g); 0.75, (h); 1.0, (i); with R=(s^4UTP)/(s^4UTP+UTP). Band 0: non thiolated molecules; bands 1 to 7: 1 to 7 s^4U residues per molecule. **b.** y axis: Wi, the relative weight of band i; x axis: number of s^4U residue per molecule; for different values of R: 2.5%, 5% and 10%.The other numbers represent the relative efficiency of incorporation of s^4U relative to U which yields the optimal fit between the experimental (full line) and theoretical distributions. Cross-hatched regions express the differences between these two distributions.

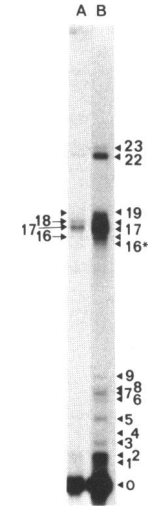

Figure 3. Separation of crosslinked molecules on 12% polyacrylamide gel. Lanes A and B are two exposures of the same sample. Band 0: uncrosslinked molecules; bands 1 to 23: crosslinked species. The% of material present in each band is: 1+2 = 6.3%; 3 = 1%; 4 =0.7%; 5 = 1%; 6+7+8 = 2%; 9 = 0.63%; 16* = 1.9%; 16 = 3%; 17 = 5.6%; 18 = 5%; 19 = 3%; 22 = 2.6%; 23 = 0.8%.

Separation of crosslinked molecules

It was known that naked fragments prepared on denaturing gels fold back to a 'native structure' in appropriate conditions (10). Therefore the monothiolated molecules eluted from a preparative APM gel were resuspended in renaturation buffer at a concentration of 0.6 pmole/μl, and irradiated immediately at 365 nm; RNA was then precipitated, resuspended in formamide and run on a 12% denaturing gel at 50°C in which crosslinked molecules are separated on the basis of loop size (see methods). Twenty three bands, numbered 1−23, migrating above band 0, the non crosslinked molecule, were detectable (Fig. 3) and of these 12 were analysed. The global yield of detectable crosslinks is 34% as deduced from the ratio radioactivity of bands 1 to 23 to total radioactivity. The yield of each band is indicated in the legend of Fig.3.The maximum yield corresponding to a given U position is 1/41 = 2.4% and accordingly some bands necessarily correspond to the reactivity of s^4U in different positions. On preparative gels several bands 1+2, 6+7+8, were eluted together because they were systematically cross contaminated and were repurified if necessary as discussed later. Crosslinked dimers which migrate in the upper part of the gel cannot interfere and will not be further considered.

Analysis of crosslinked species by reverse transcriptase primer extension and limited alkaline hydrolysis

In the first step of the analysis, the crosslinked species fractionated on 12% preparative gels were analysed by reverse transcriptase extension of primer 1 (closest to the 3′ end side of the RNA). This allows determination of the nucleotide at the 3′ side of the crosslinks present in one band eluted from the preparative gel (Fig.3). The first nucleotide incorporated corresponds to A753 (Fig.1) and any reactivity in the first helix cannot be detected. Lanes of primer extended on noncrosslinked molecules were used to control the background level of spontaneous detachment of the reverse transcriptase. Dideoxy sequencing lanes are run in parallel to localise the stops. Any crosslinking event detected by a detachment of reverse transcriptase expresses a single event per monothiolated molecule and two stops in a band correspond to one event in two molecules. Analysis of the nucleotides involved at the 3′ side of the crosslinks by extension of the primer 1 is exemplified in Fig.4. Each lane shows multiple stops corresponding to several crosslinks. In a second step we determined the nucleotide at the 5′ side of the crosslinks using two methods: extension of a second primer hybridizing inside the loop close to the 3′ side and limited alkaline hydrolysis of 5′ end labelled molecules. Analysis by elongation of primer 2 (Fig.5) gives the stops corresponding to the 5′ sides of the

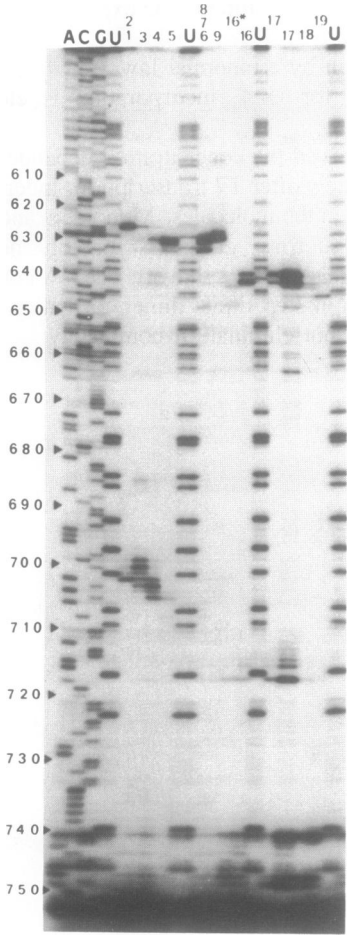

Figure 4. Reverse transcriptase extension analysis of primer 1 on crosslinked RNA species. Sequencing lanes A, C, G, U are extensions with ddT, ddG, ddC, ddA respectively. Other lanes are elongations of crosslinked RNA species from a preparative gel (see Fig.3); lanes 17 are two loads of the same sample.

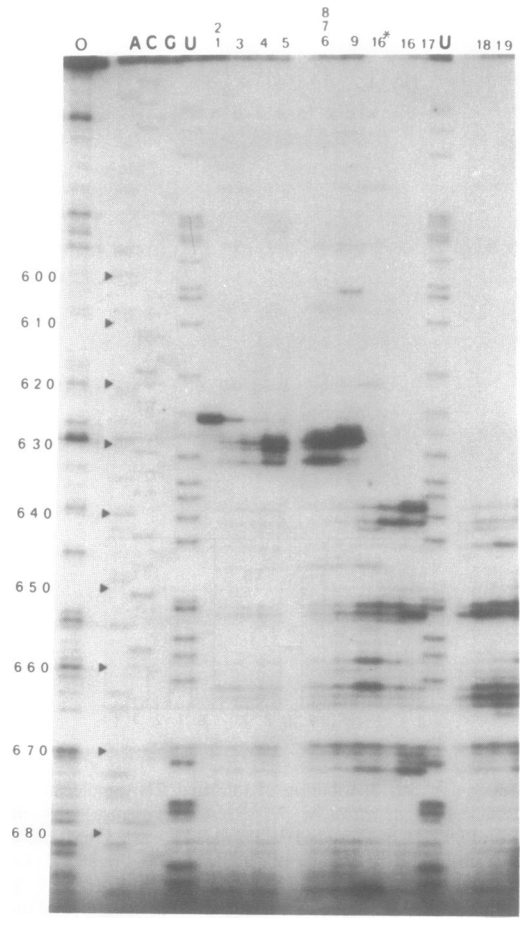

Figure 5. Reverse transcriptase extension analysis of primer 2 on crosslinked RNA species. Lane 0: control extension with non crosslinked RNA. Sequencing lanes A, C, G, U are extensions with ddT, ddG, ddC, ddA respectively. Other lanes are elongations of crosslinked RNA species from a preparative gel (see Fig.3).

crosslinks and eventually the 3′ side of crosslinks situated in helix **H** already determined with primer 1. Species 1 to 9, corresponding to short loops, cannot always be resolved by a second primer extension because the crosslink stabilises base pairing preventing its hybridization. In these cases the 5′ side of the crosslink was localised by 5′ end labelling and limited alkaline hydrolysis after repurification. Control lanes of non crosslinked molecules specifically hydrolysed by RNases T1 and U2 or statistically hydrolysed in alkaline medium were used to localise the G and A positions. A jump occurs one nucleotide before that involved in the bridge and shows a window in the ladder. The size of the window and upper part of the lane corresponds to the migration of shortened-looped molecules or molecules opened in the loop. From a band with several bridges one can read the base closest to the 5′ end provided the second species is not dominant and contributes weakly to the background in the window. Even with a pure crosslinked species, the window is not free from a background ladder; we attribute this to instability of the bridge under conditions of alkaline hydrolysis (Fig.6). For bands 16 to 19 difficulties of identification due to contaminations were solved only by separation on non denaturing gel.

Separation on non denaturing gel of reverse transcriptase elongated crosslinked species

The cross contamination of 16*, 16, 17, 18, 19 led us to devise a method allowing separation of species with similar loop size but differing in loop position. After hybridization with primer 1 and elongation with reverse transcriptase, the DNA-RNA hybrids are separated on non denaturing gels, at 4 to 10°C without urea, (Fig.7). The separation is based on the mass and charge of the added DNA and the length and rigidity of the DNA-RNA stem. The full length DNA-RNA molecule is the fastest species, spontaneous detachment of the reverse transcriptase on control molecules results in several retarded bands. Elongated species derived from bands 16 to 19 yield fractions a,b,c,d on short run gel, Fig.7a. Their resolution into subspecies on a long run gel is presented: Fig.7b, b1 to b3, band 16, d1 to d4, band 17 and Fig.7c, 4 to 8, band 18/19. The localisation of the stop at the 3′ side of the bridge is obtained by eluting the bands from the non denaturing gel and running part of each of them on sequencing gels(Fig.8). For localisation of the partner, the rest of each of the eluted bands is hybridized with primer 2 and elongated a second time with reverse transcriptase using a 5 to 10 fold higher specific activity together with appropriate controls. It should be noticed that after long run electrophoretic separation, a number of bands corresponding to the same reverse transcriptase detachment site can be obtained that may correspond to different conformers of the DNA-RNA hybrids (for example see the separation of species contained in bands 18/19: they migrate in position c (Fig.7a), are resolved in bands 4 to 8 (Fig.7c) and their analysis on Fig.8c, 8d give identical results—G741 and A663, G664, A665—for 4 and 5 and—A747, G748 and U653, G654—for 6 and 7).

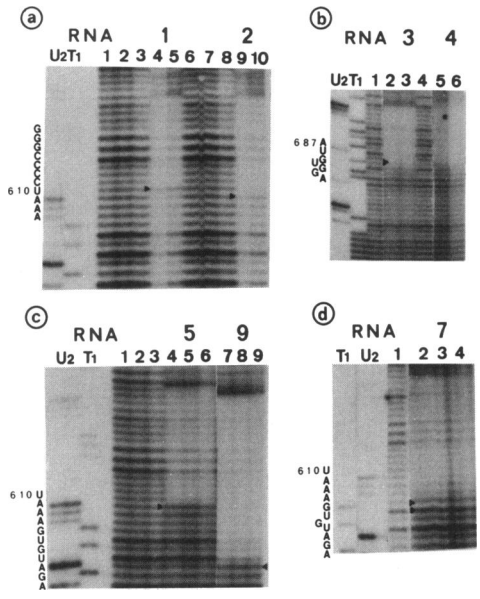

Figure 6. Alkaline hydrolysis analysis of 5′ end labeled crosslinked species. **a.** Noncrosslinked RNA: RNase U2 digest; RNase T1 digest; lanes 1, 2, 3 and lanes 6, 7, 8: 5, 10, 15 min alkaline hydrolysis. Crosslinked RNA 1, lanes 4, 5; crosslinked RNA 2, lanes 9, 10: 5, 10 min alkaline hydrolysis. **b.** Noncrosslinked RNA: RNase U2 digest; RNase T1 digest; lanes 1, 5: 5 min alkaline hydrolysis. Crosslinked RNA 3, lanes 3, 4; crosslinked RNA 4, lanes 5, 6: 5, 10 min alkaline hydrolysis. **c.** Noncrosslinked RNA: RNase U2 digest; RNase T1 digest; lanes 1, 2, 3: 5, 10, 15 min alkaline hydrolysis. Crosslinked RNA 5, lanes 4, 5, 6; crosslinked RNA 9, lanes 7, 8, 9: 5, 10, 15 min alkaline hydrolysis. **d.** Noncrosslinked RNA: RNase T1 digest; RNase U2 digest; lane 1: 15 min alkaline hydrolysis. Crosslinked RNA 7, lanes 2, 3, 4: 5, 10, 15 min alkaline hydrolysis.

Identified crosslinks

The crosslinks which have been identified are summarized in Table I. Only a few detailed examples will be described.

Bands 1+2 (Fig.3) have two major stops at A702 and G626 implying that U701 and U625 are involved in a crosslink. The partner of U701 was identified by use of primer 2 which gives stops at A687 (non visible on Fig.5) and G626 (Fig.5), the first bridge is therefore **U686 < > U701**. The partner of G626 was identified by limited alkaline hydrolysis of 5′ end labelled and repurified bands 1+2 (see Methods and Fig.6a). The jump in

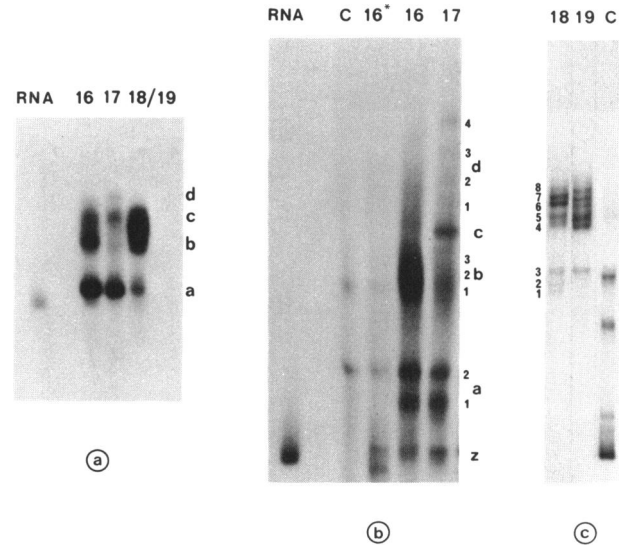

Figure 7. Non denaturing gel separation of reverse transcriptase elongated crosslinked molecules. **a.** Short run separation of species 16, 17, 18/19. **b.** Long run separation of 16*, 16, 17. **c.** Long run separation of 18/19. Lane c: elongation on non crosslinked RNA; RNA: position of non elongated RNA molecules.

lanes 4 and 5 is one nucleotide before U610 and in lanes 9 and 10, one nucleotide before C611 corresponding to bridges **U610, C611 < >U625.**

Bands 6+7+8 have very similar mobilities (Fig.3) and were not separated, the identified crosslinks (Table I) correspond to species 7 which is dominant.

Band 16* and 16 both show the same stops with primer 1, (Fig.4) at G748, A747 (stronger in 16*); the stop which follows, at G745, corresponds to a spontaneous detachment of reverse transcriptase. Primer 2 (Fig.5) reveals two characteristic stops at A663 and C660. The bridges are therefore **U659 < >A746, A747** and **U662 < >U740.**

Band 16. Two primer 1 stops are characteristic of band 16: A642 of high, and A640 of low, intensity, implying in the bridge G639 and U641; their intensities are reversed in the first lane of band 17 (Fig.4). Primer 2 gives the same stops A642 and A640 (Fig.5). On non denaturing gel these species migrate in bands a1 and a2 (Fig.7b).The partners of G639 and U641 could not be identified. The stop at A663 is not found reproducibly (absent in Fig.5 for example) and corresponds to a pyr-A susceptibility.

Band 17. The crosslinked species were completely resolved by analysis of reverse transcriptase elongated species on non denaturing gel Fig.7b (species 17c and 17d1 to d4). Species 17c corresponds to stops at G748 (primer 1, Fig.8a) and U653 (primer 2, Fig.8b), the bridge is therefore **U652 < >A747.** Species 17d1 to d4 correspond to different types of bridge: **G671, U672, A673 < >U717** and **U672 < >G713, G714, A715, U717.**

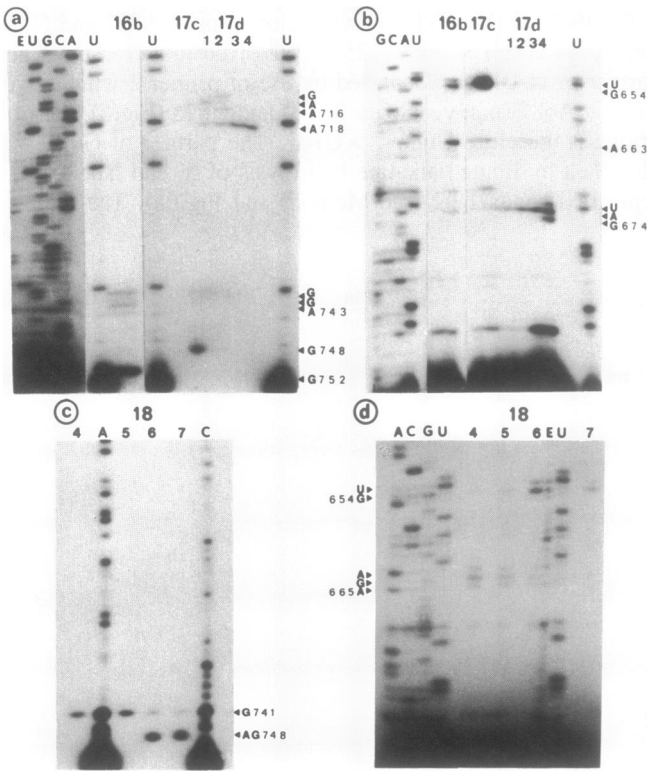

Figure 8. Analysis of reverse transcriptase elongated species separated on non denaturing gel. **a.** Stops corresponding to the first elongation (primer 1) of 16 b, 17 c and 17 d1, d2, d3, d4, (Fig.7). **b.** Stops corresponding to the second elongation (primer 2) of these same species. **c.** Stops corresponding to the first elongation (primer 1) of 18 4, 5, 6, 7 (Fig.7). **d.** Stops corresponding to the second elongation (primer 2) of the same species.

Bands 22 and 23. Elongation of primer 1 with reverse transcriptase indicates bridges involving the additional sequence located between the transcription start and the 5′ end of the insert (28 nucleotides).

DISCUSSION

Separation of crosslinked molecules

The crosslinked molecules are separated on a denaturing gel mainly on the basis of loop size which increases from 15, in bands 1+2, to 96 nucleotides, in bands 18/19. There are some exceptions however to this general behaviour (see Table I). The loop size in 3 (13 nucleotides) is shorter than in 1+2 (15 nucleotides). Also bands 17 and 18/19 have components 42−47 and 77−79 nucleotides respectively in a region where most loops

Table 1. Identified crosslinks

Samples	Crosslinks	Loop sizes
1+2	U610, C611 < >U625	15−16
	U686 < >U701	16
3	U684, U686 < >G698, C699, G700	13−17
4	U686 < >A702, A704	17−19
5	U610 < >A629, A630	20−21
6+7+8	G606, A607 < >U632	26−27
9	U603 < >G627, G628, A629	25−27
16*	U659 < >A746, A747	88−89
	U662 < >U740	79
16	U652, U653 < >U751	99−100
	n.d. < >U641	
17	n.d. < >G639	
	U652 < >A747	96
	G671, U672, A673 < >U717	45−47
	U672 < >G713, G714, A715, U717	42−46
18/19	U662, A663, G664 < >U740	77−79
	U652, U653 < >A746, A747	94−96

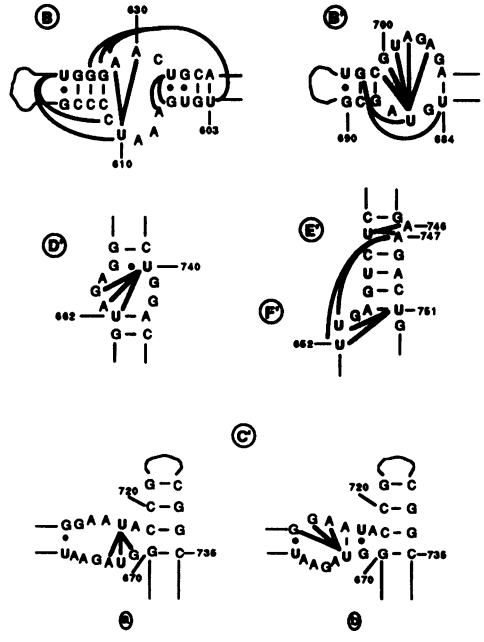

Figure 9. Summary of identified bridges and proposed secondary structures for section C′.

have sizes between 80 to 100 nucleotides. It can be noticed that the corresponding bridges are at the border of bihelical regions respectively in regions **B'**, **C'** and **D'**. These data suggest that anomalous retardation is due to the stabilisation by the bridge of the neighbouring helical region under gel electrophoresis conditions. This qualitative explanation appears also to apply to band 16 (loop size 99−100 nucleotides) which migrates slightly faster than bands 17, 18, 19, where the main components have loop sizes of 77 to 96 nucleotides.

Reverse transcriptase analysis of the bridges

The localisation of the bridges is based primarily on the reliability of the detachment of reverse transcriptase one nucleotide before the bridge. This is well documented in the case of psoralen crosslinks (22) but should be carefully reexamined here. In a number of cases several stops occur in succession and it may be asked whether this indicates several bridges or alternatively a stuttering of reverse transcriptase before a unique bridge. The following data were taken into account. In heterologous bridges joining U to A, C, G, clear stops occur one nucleotide before the Us 625, 632, 659, 717 and 740: the stop at G626 unambiguously designates U625 as the crosslinked residue. The same argument can be applied in the case of A704 in **U686<>A704**. The identical sensitivity of A746 and A747 towards chemical probes (11) strongly suggests that stops at A747 and G748 correspond to the involvement of both As in **U659<>A746, A747**. Stuttering at other successive positions cannot however be ruled out.

From the observation of complete elongation with primer 1 on purified crosslinked molecules we suspected that reverse transcriptase could pass over certain bridges with low efficiency. This was confirmed by the identification of partners of several bridges by primer 2 and by alkaline hydrolysis analysis: the partner corresponds to a second weak stop in elongations of primer 1. In the case of band 9: **U603<>A629** a stop at G604 before U603 corresponds to passage over A629 (Fig.4). However when the partner is distant in the primary structure we do not detect any secondary stop; this probably is due to the spontaneous detachment of the enzyme at each base which reduces a second low signal to an undetectable level.

Structure of the RNA fragment

The *E.coli* rRNA region extending from nucleotides 578 to 756 in domain 2 is essentially formed by two imperfect helices, **H** and **H'**, connected by a bifurcation loop **F'** as established by extensive chemical and enzymatic probing as well as directed mutagenesis (fig.1). The *in vitro* synthesized fragment folds into the same structure as 16S RNA with the exceptions of C618 and its 5' phosphate (loop **A**) and of loop **C'** which is less exposed to chemical probes in 16S RNA. It is worth noting that the naked fragment contains the complete information for binding protein S8 to helix **H** and protein S15 to helix **H'** (10, 11, 12). With respect to the secondary structure of fig.1, the intramolecular crosslinks identified here are either short-range or long-range.

Short-range crosslinks. Helix **H** is formed by three bihelical regions **1, 2, 3** connected by loops **A** and **B** and hinge region **C**. Reactivity of A640, A642 towards chemical probes has led to alternative models of **C** (10). Residues G639, U641, A642, U644 were observed to crosslink to the opposite strand, but the partner sites (which probably include U597) could not be identified. The fairly stable helices **1** and **2** are connected by loop **B** where most of the bases are fully accessible to chemical probes (10), in agreement with formation of bridges **U610<>A629, A630** and **U632<>A607, G606**.

Helix **H'** is formed by five bihelical segments **1'** to **5'** connected by loops **A'** to **D'** and hinge region **E'** (fig.1). The fortunate presence of U positions within all loops and hinge regions as well as at the extremities of minihelices allows crosslinking to occur in all regions displaying structural flexibility as it is notably the case in loop **F'**, **U751<>U652, U653** and in **E'** where U659 is able to crosslink both A746 and A747. The interior loop **D'** was proposed to be stabilised by two non-canonical G-A base pairs (11). The bridges **U740<>U662, A663, G664** show that U740 is able to come into contact with residues 662 to 664, thus indicating a much larger internal mobility than expected. This could be due to a transient rearrangement (discussed below) allowing U740 to base pair with A665.

The large asymetrical purine rich loop **C'** behaves as fully exposed in the fragment while it appears to be conformationnally heterogeneous in 16S RNA (11). Of the four U residues present within this loop at positions 672, 677, 717, 723, only two yield detectable crosslinks namely **U672<>G713, A714, A715** and **U717<>G671, U672, A673**. Extensive internal mobility within loop **C'** would have led to a much larger distribution of crosslinks, including reactivity of U723. The contrasting picture obtained by crosslinking and chemical data (11) therefore strongly suggests that **C'** is able to fold into alternative conformations leading either to the observed crosslinks or to non detectable reactivity. For example in the structures proposed in Fig.5 of (11), U672 and U717 are involved in non reactive bihelical regions unable to yield the observed bridges. A more favourable situation is encountered if the oligo C sequence 735 to 739 is moved one residue allowing C735 to base pair with G670. The rearranged region 3' is then able to form a continuous helix with three successive G-C base pairs, sequestering U723 within a small loop (Fig.9 C'). This is supported by elongations with primer 2 in which reverse transcriptase strongly detaches at G670 instead of G671. Alternative subspecies of this conformation possibly stabilised by non canonical base pairs (11) could led to the observed bridges (Fig.9 C' a,b). Residues present within loop **B'** appear to be exposed to solvent. This is in agreement with the bridge linking U686 to the opposite strand from position 698 to 702 and to A704. The crosslinks **U684<>G698, C699, G700** require a more specific structural fluctuation. Owing to the low stability of region **1'** an alternative transient pairing of U697, G, C, G700 to A687, U, G, U684 could bring the two concerned regions to close proximity.

Long-range crosslinks. Five U positions are involved in crosslinks unexpected from the secondary structure. In helix **H U603<>G627, G628, A629** and **U610, C611<>U625** connect loop **B** to stems **2** and **1** respectively. In helix **H'** the bridges **U652, U653<>A746, A747** link the two extremities of stem **5'**. The bihelical regions involved in these bridges appeared fairly stable in native conditions with the exceptions of the G606-U632 and U659-A747 pairs found respectively at the extremities of stems **2** and **5'**. It thus seems unlikely that these crosslinks occur through opening and complete rearrangement of the concerned bihelical regions. The situation prevailing here is reminiscent of the one encountered in native *E.coli* tRNA where the specific formation of the 8−13 photocrosslink is directed by the structure.

The s^4U residue at position 8 is stacked upon C13 encompassing the four base pairs of the dihydrouridine stem by means of two extended ribose-phosphate linkages (4).However in contrast to tRNA where the 3D structure is strongly stabilized by a set of tertiary interactions the long range crosslinks observed here appear to occur through transient contacts resulting mainly from considerable flexibility of loops and hinge regions and probably also of some of the terminal base pairs of minihelices.

s^4U photoreactivity

Bands 1 to 19 correspond to 32% of the monothiolated RNA and involve at least 17 U positions. The s^4U distribution within chains being random the maximal yield of crosslinked molecules is expected to be 17/41 = 41%. Hence the reactive U positions photoreact almost quantitatively (80%) with their targets. Stable covalent bonds are formed not only with pyrimidines (a single C at position 699 is involved) but also with many purine sites. Reactivity of the bases follows the decreasing order U, A, G, C in line with the data observed in mRNA-ribosome interaction (7) or after photocoupling of s^4U with nucleosides in solution (5). All bridges identified here involved partners located on opposite strands in the secondary structure since the minimal size of the loop leading to crosslink detection is in the range of 13–15 nucleotides. Hence neither reactivity at the extremities of helices **H** and **H'** nor formation of small loops within the same strand could be detected. It is remarkable that most s^4U reactive sites are found in regions allowing structural flexibility i.e. loops and bubbles. Reactivity of s^4U at the border of stable bihelical regions is in general high with the notable exception of U677. In contrast, of the 16 U positions present within stable helical regions, only two, U625 and U603, were reactive. Hence crosslinking to the opposite strand is generally prevented in a Watson–Crick double helix. This may result from: i) unfavourable orientation in spite of the possibility of breathing of small helical regions, ii) efficient quenching by stacking interactions with neighbouring residues.

ACKNOWLEDGEMENTS

We are grateful to J.L. Fourrey for s^4UTP, G.L. Igloï for APM and D.H. Hayes for careful reading of the manuscript. This work was supported by the CNRS, Paris 7 University, Pierre et Marie Curie University and the ANRS (action coordonnée no. 4).

REFERENCES

1. Brimacombe, R., Atmadja, J., Stieger, W. & Schüler, D. (1988) J. Mol. Biol. **199**,115–136.
2. Wollenzien, P.L., Murphy, R.F., Cantor, C.R., Expert-Bezançon, A. & Hayes, D.H. (1985) J. Mol. Biol. **184**, 67–80.
3. Teare, J. & Wollenzien, P.L. (1990) Nucleic Acids Res. **18**, 855–864.
4. Favre, A. (1990) Bioorganic photochemistry: photochemistry and the nucleic acids Vol. 1 H. Morrison Ed.; J. Wiley & son, Inc; pp. 379–425.
5. Favre, A., Lemaigre Dubreuil, Y., & Fourrey, J.L. (1991) New Journal of Chemistry (in press).
6. Tate, W., Greuer, B. & Brimacombe R. (1990) Nucleic Acids Res. **18**, 6537–6544.
7. Wollenzien, P., Expert-Bezançon, A., Favre, A. (1991) Biochemistry **30**, 1788–1795.
8. Kean, J.M. & Draper, D.E. (1985) Biochemistry **24**, 5052–5061.
9. Mougel, M., Ehresmann, B. & Ehresmann, C. (1986) Biochemistry **25**, 2756–2765.
10. Mougel, M., Eyermann, F., Westhof, E., Romby, P., Expert-Bezançon, A., Ebel, J.P., Ehresmann, B. & Ehresmann, C. (1987) J. Mol. Biol. **198**, 91–107.
11. Mougel, M., Philippe, C., Ebel, J.P., Ehresmann, B. & Ehresmann, C. (1988) Nucleic Acids Res. **16**, 2825–2839.
12. Gregory, R.J., Cahill, P.B.F., Thurlow, D.L. & Zimmermann, R.A. (1988) J. Mol. Biol. **204**, 295–307.
13. Prescott, C.D. & Göringer, H.U. (1990) Nucleic Acids Res. **18**, 5381–5386.
14. Moazed, D. & Noller, H.F. (1986) Cell **47**, 985–994.
15. Scheit, K.H. (1968) Chem. Ber. **101**, 1147.
16. Igloi, G.L. (1988) Biochemistry **27**, 3842–3849.
17. Silberklang, M., Gillum, A.M. & Rajbhandry, U.L. (1979) Methods in Enzymology **59**, 58–109.
18. Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989) Molecular cloning. Cold Spring Harbor Laboratory Press.
19. D'Alessio, J.M. (1982) in Gel electrophoresis of nucleic acids (Richmond, D. & Hames, B.D. eds) pp. 173–197.
20. Cramer, F., Gottschalk, E.M., Matzura, H., Scheit, K.H. & Sternbach, H. (1971) Eur. J. Biochem. **19**, 379–385.
21. Favre, A., Bezerra, R., Hajnsdorf, E., Lemaigre Dubreuil, Y. & Expert-Bezançon, A. (1986) Eur. J. Biochem. **160**, 441–447.
22. Ericson, G. & Wollenzien, P. (1988) Anal. Biochem. **174**, 215–223.