Precise determination of RNA–protein contact sites in the 50 S ribosomal subunit of *Escherichia coli*

Bernd THIEDE, Henning URLAUB[1], Helga NEUBAUER, Gerlinde GRELLE and Brigitte WITTMANN-LIEBOLD[2]
Max-Delbrück-Centrum für Molekulare Medizin, Robert-Rössle-Strasse 10, D-13125 Berlin, Germany

RNA–protein cross-linked complexes were isolated and purified to obtain precise data about RNA–protein contact sites in the 50 S ribosomal subunit of *Escherichia coli*. N-terminal microsequencing and matrix-assisted laser desorption ionization MS were used to identify the cross-linking sites at the amino acid and nucleotide levels. In this manner the following contact sites of five ribosomal proteins with the 23 S rRNA were established: Lys-67 of L2 to U-1963, Tyr-35 of L4 to U-615, Lys-97 of L21 to U-546, Lys-49 of L23 to U-139 or C-140 and Lys-71 and Lys-74 of L27 to U-2334.

INTRODUCTION

RNA–protein interactions are known to have crucial roles in gene expression, gene regulation and cell function [1,2]. Several conserved amino acid sequence motifs for RNA binding as well as secondary structure elements of RNA such as hairpins, bulges and loops for protein binding have been identified as contact sites [3–6]. However, our knowledge of the molecular structure of RNA–protein complexes is limited because only a few structures have been solved so far [6,7]. RNA–protein co-crystals, which can be difficult to generate, are required for X-ray analysis. In contrast, valuable structural information can be obtained by cross-linking techniques without having a crystal structure. Furthermore cross-linking agents can be applied to the study of multicomponent complexes such as the ribosome. A strategy has been established after cross-linking for the precise determination of RNA–protein contact sites in the 30 S ribosomal subunit of *Escherichia coli* at the molecular level. In this method, the RNA–protein cross-links were reduced to oligoribonucleotide–peptide complexes by digestion with ribonuclease and endoprotease, and after purification were analysed by N-terminal sequencing and matrix-assisted laser desorption ionization MS (MALDI-MS) [8,9]. This type of knowledge about RNA–protein contact sites in the ribosome at the molecular level is essential for an understanding of the functional implications of the constituents of the ribosome in detail [10–12]. Thus the cross-linking data obtained are required for the refinement of ribosome models [13–15] in combination with the known individual protein structures of ribosomal proteins [16]. The refined models can be inserted into maps of the overall topography of the ribosome as derived from, for example, cryoelectron microscopy images [17–20].

Here we report the determination of five RNA–protein contact sites at the molecular level within the 50 S ribosomal subunit of *E. coli*. The new cross-linking data are useful in refining the three-dimensional model of the large ribosomal subunit and enhance our knowledge of the structural details of RNA–protein recognition.

EXPERIMENTAL

Materials

2-Iminothiolane was obtained from Pierce (Rockford, IL, U.S.A.). Lys-C, Glu-C and $5' \rightarrow 3'$ phosphodiesterase (calf spleen) were purchased from Boehringer-Mannheim (Mannheim, Germany). RNAse A, ACTH (18–39) and α-cyano-4-hydroxycinnamic acid were obtained from Sigma (Deisenhofen, Germany); ribonuclease T_1 was from Calbiochem (San Diego, CA, U.S.A.), RNAsin was from Promega (Madison, WI, U.S.A.). All other chemicals were of pro analysis grade or ultra pure and were purchased from Merck (Darmstadt, Germany).

Preparation of 50 S ribosomal subunits from *E. coli*, cross-linking of ribosomal subunits, generation of cross-linked oligoribonucleotide-peptide complexes and size-exclusion chromatography were precisely performed as in Urlaub et al. [8].

Reverse-phase HPLC (RP-HPLC)

The RNA pool derived from 300 A_{260} units after cross-linking of *E. coli* 50 S subunits and after two stages of size-exclusion chromatography was digested in 1 mM EDTA with 10 μg of RNase T_1 for 2 h at 50 °C. Subsequently the endoprotease digestion was performed after adjusting the solution to 25 mM Tris/HCl, pH 7.8, and 2 mM EDTA with either 3 μg of Glu-C or 0.39 unit of Lys-C endoprotease by incubation for 16 h at 37 °C. Two aliquots of 150 A_{260} units were separately injected on a Vydac C_{18} column (250 mm × 4 mm; 300 Å; The Separation Group, Hesperia, CA, U.S.A.). A gradient of 10–45 % solvent B was applied with a flow rate of 0.5 ml/min at room temperature over a period of 240 min. As solvents, aqueous 0.1 % (v/v) trifluoroacetic acid (solvent A) and 0.085 % (v/v) trifluoroacetic acid in acetonitrile (solvent B) were used. Single fractions showing absorbance at 220 and 260 nm were used for further analysis; they were dried under vacuum and stored at −80 °C. The samples were dissolved in aqueous 0.1 % (v/v) trifluoroacetic acid/50 % (v/v) acetonitrile (1:1, v/v) for N-terminal sequence and MS analysis.

Abbreviations used: MALDI, matrix-assisted laser desorption ionization; RP-HPLC, reverse-phase HPLC.

[1] Present address: Institut für Molekularbiologie und Tumorforschung, Emil-Mannkopff-Str. 2, D-35037, Germany.
[2] To whom correspondence should be addressed (e-mail liebold@mdc-berlin.de).

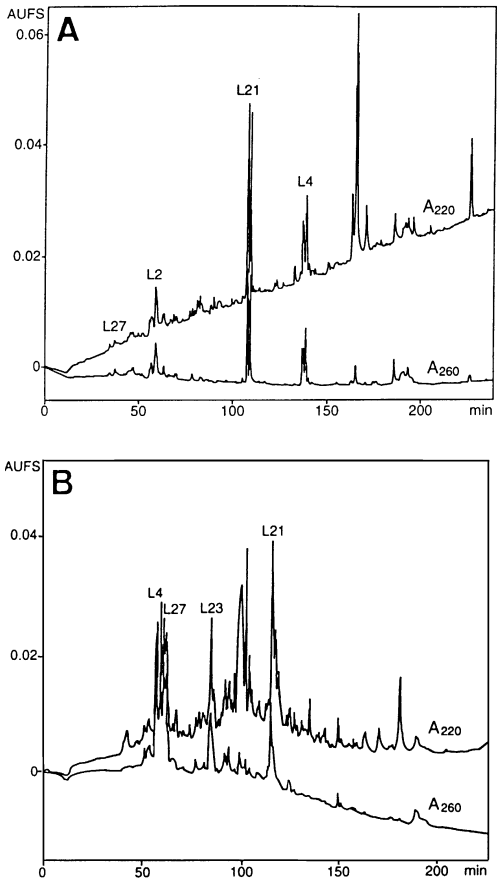


Figure 1 RP-HPLC analysis of oligoribonucleotide–peptide complexes derived from the 50 S ribosomal subunit of *E. coli* after cross-linking with 2-iminothiolane and UV irradiation

(**A**) Cross-linked 50 S subunit (150 A_{260} units) after digestion with Lys-C and ribonuclease T_1. (**B**) Cross-linked 50 S subunit (150 A_{260} units) after digestion with Glu-C and ribonuclease T_1. The proteins of the generated peptide–oligoribonucleotide complexes are indicated. Individual fractions of the HPLC peaks were used for peptide and oligoribonucleotide sequencing.

N-terminal sequence analysis

Approx. 1–5 pmol of the cross-linked oligoribonucleotide–peptide complexes dissolved in aqueous 0.1 % trifluoroacetic acid/50 % acetonitrile (1 : 1) were loaded directly on a PROCISE® protein sequencer (Applied Biosystems, Foster City, CA, U.S.A.) to identify the cross-linked peptide moieties. The cross-linked amino acids were not detected in the appropriate Edman degradation steps. The ribosomal protein corresponding to the obtained peptide sequence was derived by a comparison against the Swiss-Prot protein sequence database by using the program BLITZ.

MALDI-MS

The mass spectra of the complexes were recorded with a MALDI–time-of-flight mass spectrometer (VG TofSpec; Fisons Instruments, Manchester, U.K.; or Bruker-Reflex; Bruker-Franzen Analytik, Bremen, Germany). For the mass determination, the dried-drop method was used by mixing 0.7 μl of matrix solution [saturated solution of α-cyano-4-hydroxycinnamic acid in aqueous 0.1 % trifluoroacetic acid/50 % acetonitrile (1 : 1)] and 0.5 μl of the sample dissolved in aqueous 0.1 % trifluoroacetic acid/50 % acetonitrile (1 : 1) on the sample holder. For the UV cross-link, 0.2 μl (200 fmol) of ACTH (18–39) was added additionally as an internal standard. The spectra were obtained in the positive-ion mode at 22 kV by summing over 20–40 laser shots.

Oligoribonucleotide sequencing

The vacuum-dried oligoribonucleotide–peptide complexes were dissolved for alkaline hydrolyis in 20 μl of aqueous NH_4OH, pH 10.5, and incubated for 15 min at 95 °C. For treatment with phosphodiesterase, the dried samples were dissolved in 20 μl of 1 mM Tris/HCl, pH 7.8, and incubated for 1 h at 37 °C with 0.002 unit of 5′ → 3′ phosphodiesterase. The hydrolysed samples were then dried under vacuum and dissolved in aqueous 0.1 % trifluoroacetic acid/50 % acetonitrile (1 : 1) for the mass analysis as described above.

RESULTS

Isolation of cross-linked oligoribonucleotide–peptide complexes

The 50 S ribosomal subunits of *E. coli* were cross-linked with 2-iminothiolane followed by UV irradiation. The RNA–protein complexes were digested with ribonuclease T_1 and endoprotease Lys-C or Glu-C to obtain oligoribonucleotide–peptide complexes. Purification by size-exclusion chromatography and finally

Table 1 Contact sites between ribosomal proteins and the 23 S rRNA within the 50 S ribosomal subunit of *E. coli*

The deduced cross-linked nucleotides and amino acids are shown in bold. The mass difference was calculated by subtraction of the peptide mass from that of the corresponding total complex. An asterisk indicates that the cross-linked nucleotide sequence can be verified by comparing it with the data in [21]. Abbreviation: n.d., cross-linked amino acid not determined.

Ribosomal protein	Endoprotease	Peptide sequence	Mass difference (Da)	Nucleotide composition	Nucleotide sequence	Cross-linked nucleotide in 23 S rRNA
L2	Lys-C	^{59}QAYRIVDF**K**RNK^{70}	1608.5	AC_2UG	ACC**U**G	U-1963
L4	Glu-C	^{26}ALVHQVVVA**Y**AAGARQGTRAQKTR^{49}	1656.4	A_3UG	AA**U**AG*	U-615
	Lys-C	^{7}DAQSALTVSETTFGR^{21}				
L21	Glu-C	^{71}KVKIVKFRRRKHYRKQQGHRQWFTDV**K**IT^{99}	1609.9	ACU_2G	CU**U**AG*	U-546
	Lys-C	^{86}QQGHRQWFTDV**K**ITGI^{101}				
L23	Glu-C	^{43}IKAAVQ**K**LFE^{52}	1586.4	CU_3G	UU**UC**G*	U-139 or C-140
L27	Glu-C	^{70}VKGP**K**NRKFISIEAE^{84}	1962.4	A_3CUG	CA**U**AAG*	U-2334
	Lys-C	^{70}V**K**GPKL^{75}				

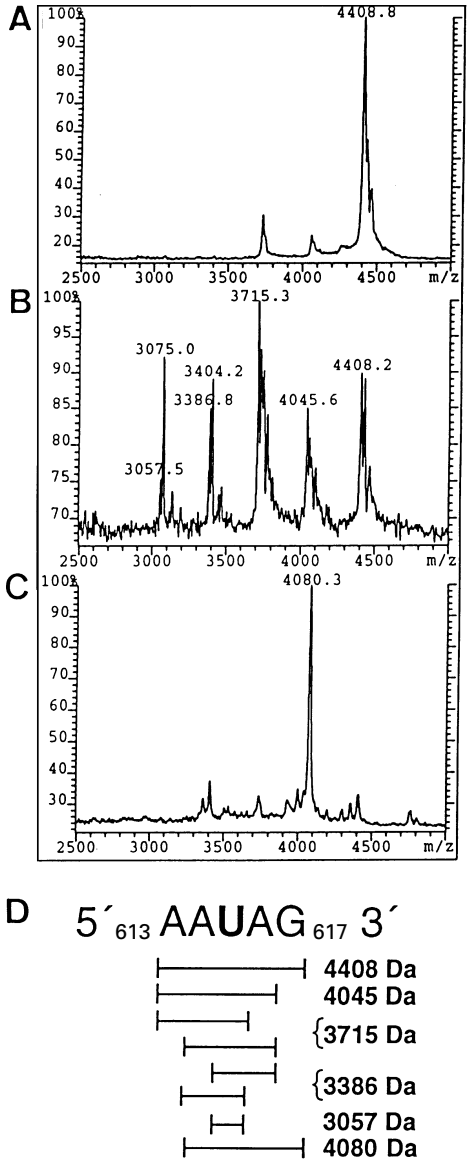


Figure 2 MALDI-MS analysis of the L4 cross-link

Mass spectrum of the total complex (**A**) and after hydrolysis with aqueous NH_4OH (**B**) and 5′ → 3′ phosphodiesterase (**C**). The deduced oligoribonucleotide moiety is shown in (**D**). Additional mass peaks of +18 Da (H_2O) were detected after hydrolysis with aqueous NH_4OH.

by RP-HPLC led to the isolation of these complexes by monitoring the absorption at 220 and 260 nm (Figure 1).

N-terminal sequence analysis

The N-terminal sequence analysis of the peptide moiety enabled the corresponding ribosomal protein to be identified: a gap in the sequence defined the cross-linking position (Table 1). The cross-linked amino acids in the ribosomal proteins L2, L4, L21 and L23 were clearly identified at a single position. Slightly different sites were obtained for L27 when different endoproteases were used. The neighbouring lysine residues at positions 71 and 74 respectively were identified as cross-linking sites to the same oligoribonucleotide sequence. This observation can be explained easily on the assumption that both lysine residues are involved in this particular RNA cross-link.

MS

The oligoribonucleotide moiety was analysed by MALDI-MS before and after partial hydrolysis (Figure 2). The nucleotides at the 5′ and 3′ ends were partly hydrolysed from the oligoribonucleotide–peptide complexes by treatment with aqueous NH_4OH; this resulted in fragments of different lengths. The additional 5′ → 3′ phosphodiesterase digest defined fragments from the 5′ end. The mass difference between the total complex and the corresponding peptide enabled the determination of the composition of the cross-linked oligoribonucleotide, whereas the oligoribonucleotide sequence and cross-linking position were revealed after the partial hydrolysis (Table 1).

The composition analysis of the 2-iminothiolane cross-links could be calculated with a mass accuracy of better than 0.7 Da by using the corresponding detected peptide mass as an internal calibration. An internal standard (ACTH 18–39) for calibration was added to the UV cross-linked complex of L4 because the corresponding peptide mass was not detected.

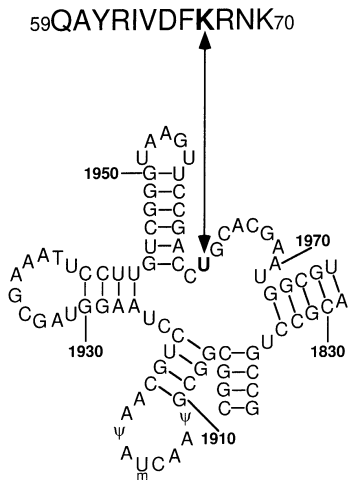
The contact sites of the ribosomal proteins L2, L4, L21 and L27 with the 23 S rRNA were localized to a single nucleotide, whereas two possible neighbouring nucleotides were determined for L23. These new results are in accordance with other RNA–protein cross-linking data from the 50 S ribosome [21], where the entire nucleotide stretches were determined, but not the precise cross-linked nucleotide as in our studies.

DISCUSSION

The knowledge of RNA–protein contact sites at high resolution is important for combination with the information obtained from three-dimensional structures of ribosomal proteins [22–28] and RNA subdomains [11,29]. The new results in this study can be correlated with the cryoelectron microscopy images of the ribosome [17,18]. They further help to fit the known three-dimensional protein structures into the refined rRNA molecular models by fixing points [13–15,19,20,30]. Although RNA–protein interactions in the ribosomes have been studied extensively, mainly by cross-linking [14] and footprinting experiments [31], these methods fail to define the relative arrangements of the components and the precise localization of the contact sites of individual nucleotides and amino acids. For this purpose we developed an approach to determine the contact sites after cross-linking by either direct UV irradiation (zero-length cross-linker) or treatment with 2-iminothiolane (7 Å) followed by UV irradiation [8,9]. In this paper the application of this approach was demonstrated for the 50 S ribosomal subunit.

The cross-linked amino acids of the ribosomal proteins L2, L4, L21, L23 and L27 were identified. Unfortunately, the three-dimensional structures of these proteins have not yet been solved. If these structures become available, the RNA–protein interaction data could be used to fit these structures into the model of the 50 S ribosomal subunit as shown for S7 and S8 for the 30 S subunit, for example [20]. RNA-binding motifs [3,4] were not found within these proteins. Hairpin loops (L4, L21 and L23) and internal loops (L2 and L27) were identified as secondary structure motifs for protein recognition by the 23 S rRNA with our cross-link data.

The ribosomal proteins L2, L3 and L4 are the most likely candidates for an involvement in the peptidyltransferase activity;



Scheme 1 Presentation of the cross-linking site between the ribosomal protein L2 (Lys-67) and the 23 S rRNA (U-1963) in *E. coli*

The secondary structure of the 23 S rRNA refers to Brimacombe [14]. The cross-linking site within the peptide and the rRNA is connected with an arrow and is shown in bold.

L27 has an important role in ribosomal function and assembly [11]. Contact sites to the 23 S rRNA of three of these proteins were identified. The functional implication of ribosomal protein L2 is of particular interest [32]. The cross-link of L2 was located in domain IV (Scheme 1) of the 23 S rRNA, which is in accordance with nuclease protection experiments [33] in which two nucleotide regions were identified as L2-binding sites. Furthermore another cross-link of L2 was localized within the same domain of the 23 S rRNA [34].

This work was supported by grants to B.W.-L. from the Deutsche Forschungsgemeinschaft (SFB 344, YE6 and Wi 358/12-1).

REFERENCES

1 Nagai, K. and Mattaj, I. W. (1994) RNA–Protein Interactions, IRL Press, Oxford
2 Siomi, H. and Dreyfuss, G. (1997) Curr. Opin. Genet. Dev. **7**, 345–353
3 Mattaj, I. W. (1993) Cell **73**, 837–840
4 Burd, C. G. and Dreyfuss, G. (1994) Science **265**, 615–621
5 Draper, D. E. (1995) Annu. Rev. Biochem. **64**, 593–620
6 Nagai, K. (1996) Curr. Opin. Struct. Biol. **6**, 53–61
7 Nissen, P., Kjeldgaard, M., Thirup, S., Polekhina, G., Reshetnikova, L., Clark, B. F. C. and Nyborg, J. (1995) Science **270**, 1464–1472
8 Urlaub, H., Kruft, V., Bischof, O., Müller, E.-C. and Wittmann-Liebold, B. (1995) EMBO J. **14**, 4578–4588
9 Urlaub, H., Thiede, B., Müller, E.-C., Brimacombe, R. and Wittmann-Liebold, B. (1997) J. Biol. Chem. **272**, 14547–14555
10 Wool, I. G. (1996) Trends Biochem. Sci. **21**, 164–165
11 Green, R. and Noller, H. F. (1997) Annu. Rev. Biochem. **66**, 679–716
12 Siegel, V. (1997) Cell **90**, 5–8
13 Malhotra, A. and Harvey, S. C. (1994) J. Mol. Biol. **240**, 308–340
14 Brimacombe, R. (1995) Eur. J. Biochem. **230**, 365–383
15 Fink, D. L., Chen, R. O., Noller, H. F. and Altman, R. B. (1996) RNA **2**, 851–866
16 Ramakrishnan, V., Davies, C., Gerchman, S. E., Golden, B. L., Hoffmann, D. W., Jaishree, T. N., Kycia, J. H., Porter, S. and White, S. W. (1995) Biochem. Cell. Biol. **73**, 969–977
17 Frank, J., Zhu, J., Penczek, P., Li, Y., Srivastava, S., Verschoor, A., Rademacher, M., Grassucci, R., Lata, R. K. and Agrawal, R. K. (1995) Nature (London) **376**, 441–444
18 Stark, H., Mueller, F., Orlova, E. V., Schatz, M., Dube, P., Erdemir, T., Zemlin, F., Brimacombe, R. and van Heel, M. (1995) Structure **3**, 815–821
19 Mueller, F. and Brimacombe, R. (1997) J. Mol. Biol. **271**, 524–544
20 Mueller, F. and Brimacombe, R. (1997) J. Mol. Biol. **271**, 545–565
21 Osswald, M., Greuer, B. and Brimacombe, R. (1990) Nucleic Acids Res. **18**, 6755–6760
22 Liljas, A. and Garber, M. (1995) Curr. Opin. Struct. Biol. **5**, 721–727
23 Davies, C., White, S. W. and Ramankrishnan, V. (1996) Structure **4**, 55–66
24 Davies, C., Ramankrishnan, V. and White, S. W. (1996) Structure **4**, 1093–1104
25 Nikonov, S., Nevskaya, N., Eliseikina, I., Fomenkova, N., Nikulin, A., Ossina, N., Garber, M., Jonsson, B.-H., Briand, C., Al-Karadaghi, S. et al. (1996) EMBO J. **15**, 1350–1359
26 Berglund, H., Rak, A., Serganov, A., Garber, M. and Härd, T. (1997) Nature Struct. Biol. **4**, 20–23
27 Hosaka, H., Nakagawa, A., Tanaka, I., Harada, N., Sano, K., Kimura, M., Yao, M. and Wakatsuki, S. (1997) Structure **5**, 1199–1208
28 Wimberly, B. T., White, S. W. and Ramakrishnan, V. (1997) Structure **5**, 1187–1198
29 Uhlenbeck, O. C., Pardi, A. and Feigon, J. (1997) Cell **90**, 833–840
30 Mitchell, P., Osswald, M., Schueler, D. and Brimacombe, R. (1990) Nucleic Acids Res. **18**, 4323–4333
31 Noller, H. F., Green, R., Heilek, G., Hoffarth, V., Hüttenhofer, A., Joseph, S., Lee, I., Lieberman, K., Mankin, A., Merryman, C. et al. (1995) Biochem. Cell Biol. **73**, 997–1009
32 Ühlein, M., Weglöhner, W., Urlaub, H. and Wittmann-Liebold, B. (1998) Biochem. J., **331**, 423–430
33 Beauclerk, A. A. D. and Cundliffe, E. (1988) EMBO J. **7**, 3589–3594
34 Gulle, H., Hoppe, E., Osswald, M., Greuer, B., Brimacombe, R. and Stöffler, G. (1988) Nucleic Acids Res. **16**, 815–832

Received 23 February 1998/20 May 1998; accepted 11 June 1998