Preliminary NMR studies of *Thermus thermophilus* ribosomal protein S19 overproduced in *Escherichia coli*

Natalia L. Davydova[a,b], Alexey V. Rak[a], Olga I. Gryaznova[a], Anders Liljas[b], Bengt-Harald Jonsson[c], Helena Berglund[d], Torleif Härd[d], Maria B. Garber[a,*]

[a] *Institute of Protein Research, Russian Academy of Sciences, 142292 Pushchino, Moscow Region, Russia*
[b] *Molecular Biophysics, University of Lund, Chemical Center, P.O. Box 124, S-221 00 Lund, Sweden*
[c] *Department of Biochemistry, Umeå University, S-901 87 Umeå, Sweden*
[d] *Department of Biochemistry and Biotechnology at the Center for Structural Biochemistry, Royal Institute of Technology, Novum, S-141 57 Huddinge, Sweden*

Received 22 July 1997; revised version received 22 August 1997; accepted 25 August 1997

Abstract The gene for the ribosomal protein S19 from *Thermus thermophilus* was cloned, sequenced and overexpressed in *Escherichia coli*. A simple procedure for isolating the recombinant protein was developed. Preliminary NMR studies revealed a high content of α-helical secondary structure in the protein.

Key words: Ribosomal protein S19; Overexpression; Purification; Nuclear magnetic resonance; *Thermus thermophilus*

1. Introduction

Protein synthesis on the ribosome is a fundamental process in all living cells. Molecular mechanisms of translation cannot be understood without a detailed knowledge of the structure of ribosomal components. Unfortunately, only few of them could be isolated from ribosomes without denaturation in amounts sufficient for structural investigations [1]. This problem was overcome by using the methods of modern recombinant DNA technology [2,3]. Current structural studies are focusing on ribosomal proteins from thermophilic bacteria which have been overproduced in *Escherichia coli* [4].

In this work we focused our attention on the ribosomal protein S19 (rpS19) from the extreme thermophile *Thermus thermophilus*. In *E. coli*, S19 is one of the four ribosomal proteins (S3, S7, S14, S19) that interact with the 3′ major domain of 16S rRNA [5]. It has been shown that this structural domain is essential for high-affinity tetracycline binding [6]. S19 strongly protects two distinct regions of the 3′ domain: 955–1050 and the 1320 stem-loop [7]. Neutron diffraction data suggest that S19 and S14 are very closely spaced [8]. Crosslinking data show that S13–S19 is one of the main protein pairs formed in the 30S ribosomal subunit from both *Bacillus stearothermophilus* and *E. coli* [9]. Here we report the cloning, sequencing and overexpression of the gene for rpS19 from *T. thermophilus* in *E. coli* as well as the purification of the recombinant protein and its preliminary NMR analysis.

2. Materials and methods

2.1. Bacterial strains and growth conditions

T. thermophilus was grown as described by Zheltonosova et al. [10]. *E. coli* strains TG1 and BL21(DE3) were grown in SOB [12] and made competent by the method of Hanahan [11]. *E. coli* strain CJ236 (used for preparation of single-stranded DNA) and the transformed cells were grown in 2×LB [12].

2.2. Recombinant DNA techniques

Chromosomal DNA was isolated from *T. thermophilus* strain VK1 following standard protocols [13]. The uracil-containing single-stranded form of the vector was isolated from *E. coli* strain CJ236 (dut^-, ung^-) by the method of Kunkel et al. [14].

Oligonucleotide primers: S19start (5′ phosphorylated) 5′-GTTTAACTTTAAGAAGGAGATATACCATGCCGCGTAGTCTCAAGAAAGGC-3′; S19stop 5′-CAGTGAAAGTGAAACTGAATCAATCTACTTCTTCTTGGTGGCCTTGG-3′.

DNA sequencing was performed with the M13 dideoxy method [15] using Sequenase Version 2.0 Kit (U.S. Biochemicals). Universal pUC/M13 forward primer was used for M13 sequencing. Primers complementary to the expression vector were used for control sequencing of the cloning gene region: vector primer (forward) 5′-CAACGGTTTCCCTCTAGA-3′ and reverse primer 5′-TCTGAACCTTGTCCATC-3′.

Sequence compressions were solved using Taq Track Sequence Deaza System (Promega). The protein coding region after cloning into the expression vector was checked by direct sequencing from the plasmid DNA with Sequenase Version 2.0.

2.3. Cloning, overexpression and sequence analysis

Plasmid pBR322-tthS8 [16] for initial gene sequencing was kindly provided by Dr. V. Vysotskaya. The ribosomal protein S19 was cloned into pACA vector using the 'sticky feet'-directed mutagenesis protocol [17,18]. The plasmid, called pTthS19, isolated from the clone with the correct gene sequence was transformed into *E. coli* strain B834(DE3). Protein production was induced by adding IPTG to a final concentration of 0.5 mM.

The sequences for several proteins homologous to S19 were obtained from EMBL and Swiss-Prot databases. WISCONSIN software package Version 8.0 1UNIX 1994 was used for DNA sequence analysis. Amino acid sequences were aligned with the PileUp program of the GCG package with a gap penalty of 3.0.

2.4. Purification procedure

The purification procedure for the ribosomal protein S19 was almost the same as for the other recently overexpressed ribosomal proteins from *T. thermophilus* [16,17] with some modifications. Briefly, 3 g of wet cells were ground with alumina and resuspended in 100 mM Tris-HCl, pH 8.1, with 800 mM NaCl, 150 mM $MgCl_2$, 5 mM β-mercaptoethanol and DNase (0.5 mg/ml final concentration). After removal of the cell debris the supernatant was heated at 65°C for 20 min. The precipitated proteins were spun down, the supernatant was diluted to 300 mM NaCl with 50 mM Na-acetate buffer, pH 5.2, and loaded on a CM-Sepharose column at a flow rate of 1 ml/min.

*Corresponding author. Fax: (7) (95) 924 0493.
E-mail: garber@sun.ipr.serpukhov.su

PII S0014-5793(97)01112-5

```
Tth  .....MPRSL KKGVFVDDH. .LLEKVLEL. .......NAK GEKRLIKTWS  35
Bst  ......GRSL KKGPFCDEH. .LMKKIEKL. .......NET GQKQVIKTWS  34
Eco  ......PRSL KKGPFIDLH. .LLKKVEKA. .......VES GDKKPLRTWS  34
Tma  .....MGRSR KKGPYVDRK. .LLEKIRKL. .......NET GEKKVIKTWS  35

Psa  .....MTRSR KKNPFVANH. .LLKKIKKL. .......NTK GEKAIIKTWS  35
Zma  .....VTR.K KTNPFVARH. .LLAKIEKV. .......NMK EEKEIIVTWS  34
Phy  ....MPRRSI WKGSFVDAF. .LLRMKKKR. .......DLL FNRKI...WS  33
Obe  ....MPRRSI WKGSFVDAF. .LLRMKKKR. .......DLL LNRKI...WS  33

Ham  ...MNDTVTI RTRKFMTNR. .LLQRKQMVI DVLHPGKATV PKTEIREKLA  45
Xla  ...MNDTVTI RTRKFMTNR. .LLQRKQMVI DVLHPGKATV PKTEIREKLA  45
Dme  .MPGVTVKDI DQHAVTKAVA VFLKKTGKLK VPDQMDIIKT AKFKELAPYD  49
Rat  .MPGVTVKDV NQQEFVRALA AFLKKSGKLK VPEWVDTVKL AKHKELAPYD  49
Asu  MVKATSVKDV DQHEIVQHIA KFLKKSGKVK VPDWSDVTKM GISKELAPLN  50
              ++   +  ::      :: :  :                 +:   +++

Tth  RRSTIVPEMV ...GHTIAVY NGKQHVPVYI TENM..... .VGHKLGEFA  75
Bst  RRSTIFPQFV ...GHTIAVY DGRRHVPVYI TEDM..... .VGHKLGEFA  74
Eco  RRSTIFPNMI ...GLTIAVH NGRQHVPVFV TDEM..... .VGHKLGEFA  74
Tma  RASMIIPKWV ...GHGIAVY NGMKHIPVYI TENM..... .IGHRLGEFA  75

Psa  RKSTIIPIMI ...GHTIAIH NGKEHLPVYI TDRM..... .VGHKLGEFS  75
Zma  RASSILPAMV ...GHTIAIH NGKEHIPIYI TNPM..... .VGRKLGEFV  74
Phy  RRSSISPEFV ...DCSVRIY NGKTPVRCKI TEGK..... .VGHKFGEFA  73
Obe  RRSSILPEFV ...NCSVRIY NGKTPVRCKI TEEK..... .VGHKF.EFA  72

Ham  KMYKTTPDVI FVFGFRTHFG GGKTTGFGMI YDSLDYAKKN EPKHRLARHG  95
Xla  KMYKTTPDVI FVFGFRTHFG GGKTTGFGMI YDSLDYAKKN EPKHRLAKHG  95
Dme  PDWFYVRCAS ILRHLYHRSP AGVG..... ........SI TKIYGGRKRN  85
Rat  ENWFYTRAAS TARHLYLRGG AGVG..... ........SM TKIYGGRQRN  85
Asu  SDWYYVRTAS IARRLYVRSP TALMH.GLSM VAANDVAWSL IILPRLPAR.  98
     +++ + :       :       +:    +  : +           ++:++++++

Tth  PTRTYRGHGK EAKATKKK.. ......... ......... .........  93
Bst  PTATFRGHAG DDKKTKR... ......... ......... .........  91
Eco  PTRTYRGHAA DKKAKKK... ......... ......... .........  91
Tma  PTRRFGGHA. DKKAKKGELK K........ ......... .........  95

Psa  PTLNFGGF.. AKNDNKSRR. ......... ......... .........  92
Zma  PTRHFTSYES TRKDTKSRR. ......... ......... .........  93
Phy  STRKRRPSRT NIGPGRKRGK K........ ......... .........  94
Obe  FTRKRRRSRT NIGSGRKRGK K........ ......... .........  93

Ham  LYEKKKTSRK QRKERKNRMK KVRGTAKANV GAGKKPKE.. ..........  133
Xla  LYEKKKTSRK QRKERKNRMK KVRGTAKANV GAGKKKD... ..........  132
Dme  GVHPSHFCRA ADGAARKALQ ALEHARLVEK HPDGGRKLSS IGQRDLDRIA  135
Rat  GVRPSHFSRG SKSVARRVLQ ALEGLKMVEK DQDGGRKLTP QGQRDLDRIA  135
Asu  SVKPCRHSRQ SNGC..RNIQ MVTDAYL... .........P NGRKDLDRIA  134
     +++             :
```

Fig. 1. Multiple alignment of 13 amino acid sequences of the ribosomal protein S19. The first four sequences are eubacterial, followed by four from plant organelle genomes and the other five eukaryotic. Colons (:) indicate well conserved amino acids between all groups, plus (+) indicates conservation between eubacterial and plant organelle groups (bold letters). Abbreviations, accession numbers and references are: Tth, *Thermus thermophilus*, X84059 (present work); Bst, *Bacillus stearothermophilus* [21]; Eco, *Escherichia coli* [22]; Tma, *Thermotoga maritima*, Z21677 [23]; Psa, *Pisum sativum*, X59015 [24]; Zma, *Zea mays*, Y00141 [25]; Phy, *Petunia hybrida*, X57283 [26]; Obe, *Oenothera berteriana*, X61030; Ham, Syrian hamster, X52658 [27]; Xla, *Xenopus laevis*, V01443 [28]; Dme, *Drosophila melanogaster*, X73153 [29]; Rat, *Rattus rattus*, X51707 [30]; Asu, *Ascaris suum*, X75543 [31].

The recombinant protein was eluted from the column with a linear gradient of 0.3–1.0 M NaCl. The gradient fractions were analyzed by SDS-PAGE and those containing pure rpS19 were pooled and precipitated with $(NH_4)_2SO_4$. Precipitated protein was dissolved in 50 mM Na-acetate buffer (pH 6.0) with 300 mM NaCl and CM-Sepharose chromatography was repeated with the same conditions but at pH 6.0.

2.5. Mild proteolytic digestion

Digestion of the intact protein was performed with trypsin (TPCK-treated, Worthington) in an enzyme/substrate ration of 1:200 on ice and at 37°C in 20 mM Tris-HCl, pH 8.0. The sizes of the produced protein fragments were analyzed by SDS-PAGE with molecular weight markers, MW range 2512–16 949 (Pharmacia).

2.6. NMR searching

NMR was measured at 20°C on a Varian Unity 500 MHz spectrometer. NMR samples were prepared with an approximate concentration of 1 mM protein in a buffer containing 200 mM NaCl and 1 mM DTT at pH 6.0. The one-dimensional spectrum in Fig. 2 was recorded with a spectral width of 8000 Hz using 16 time domain transients and was processed with a shifted Gaussian apodization function. A two-dimensional NOESY spectrum [20] was recorded in hypercomplex mode with 8000 Hz sweep widths in both dimensions, 256 complex t1 increments and 128 transients per increment. The cross-relaxation mixing time was 150 ms. The spectrum was processed with shifted sine bell apodization functions. A simple spline function baseline correction of the acquisition domain was applied before processing of the indirectly detected dimension.

3. Results and discussion

The rpS19 gene from *T. thermophilus* encodes a polypeptide chain of 93 amino acids (10 569). The complete nucleotide sequence has been deposited in the EMBL nucleotide sequence database (accession number X84059). The comparison of the amino acid sequences of rpS19 from *E. coli*, *B. stearothermophilus* and *T. thermophilus* shows 67% identity between *E. coli* and *T. thermophilus*, and 68% identity between *T. thermophilus* and *B. stearothermophilus*. Unlike rpS19 from *E. coli* and *B. stearothermophilus*, in rpS19 from *T. thermophilus* the initial Met is not processed. No significant variations can be noted within the main amino acid groups: acidic, basic, hydrophobic, aromatic. It should be mentioned that the Cys residue is absent in *T. thermophilus* rpS19.

The alignment of the rpS19 amino acid sequence from *T. thermophilus* with three eubacterial, four plant organelle and two eukaryotic sequences of homologous proteins (Fig. 1) shows that eubacterial and plant organelle sequences are very similar to each other but differ from the eukaryotic protein. There are only two amino acid residues (16 and 54, Tth numbering) which are almost conserved throughout all the organisms. The conservation of basic residues close to the N- and C-termini is characteristic for eubacterial and plant organelle sequences (3, 7, 18, 21, 28, 29, 36, 37, 70, 78, 88 and 91, Tth numbering). Some of the residues are highly conserved throughout all the sequences.

There are two highly conserved regions in eubacterial and plant organelle sequences (33–42 and 66–78, Tth numbering) with His^{68} conserved even in some eukaryotes. This amino acid residue was shown to be involved in crosslinking of pro-

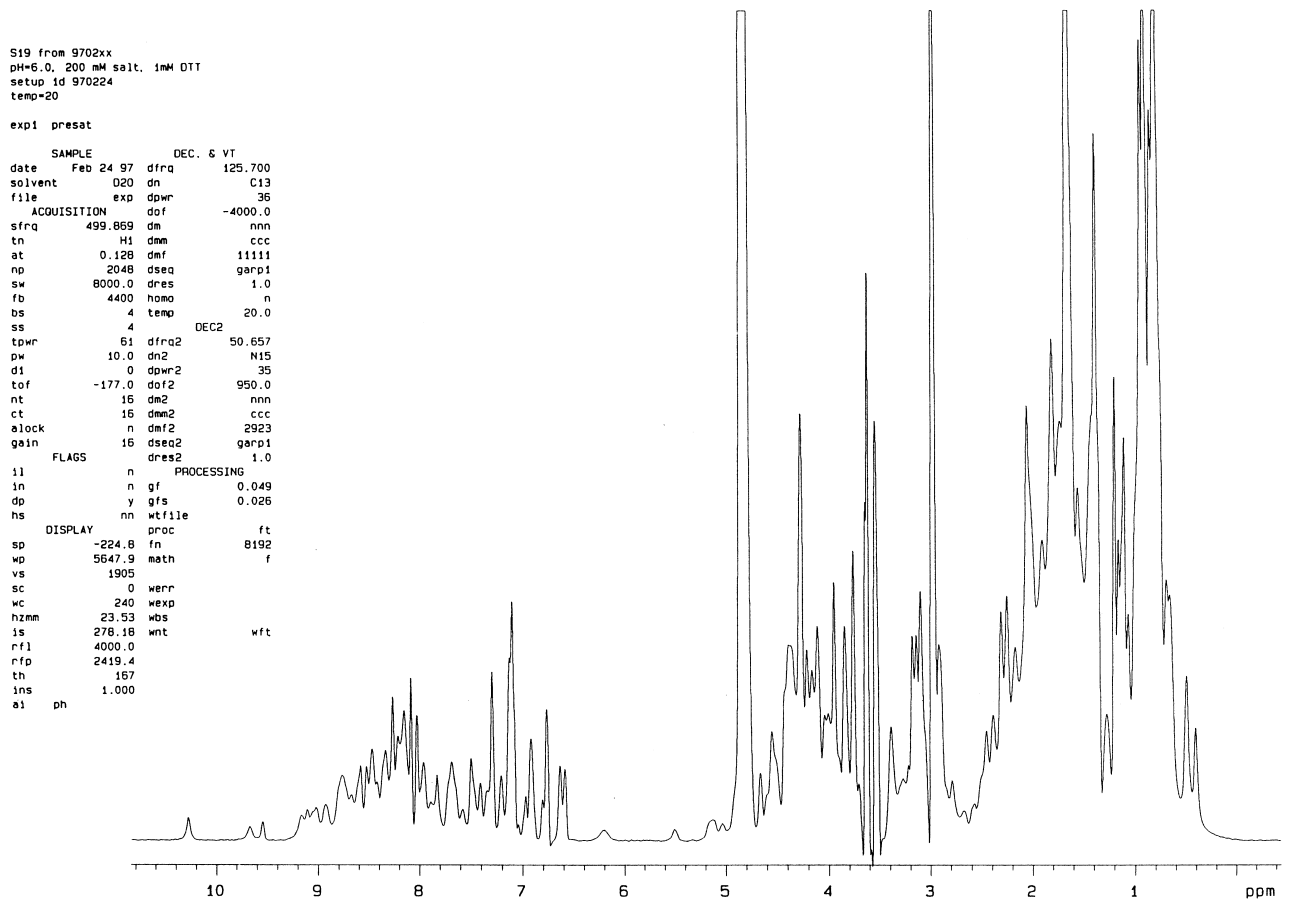

Fig. 2. One-dimensional ^1H spectrum of ribosomal protein S19 at 500 MHz. The spectrum is typical of a folded non-aggregated protein. The sharper resonances around 3.0 and 3.5 ppm are due to the presence of DTT in the sample.

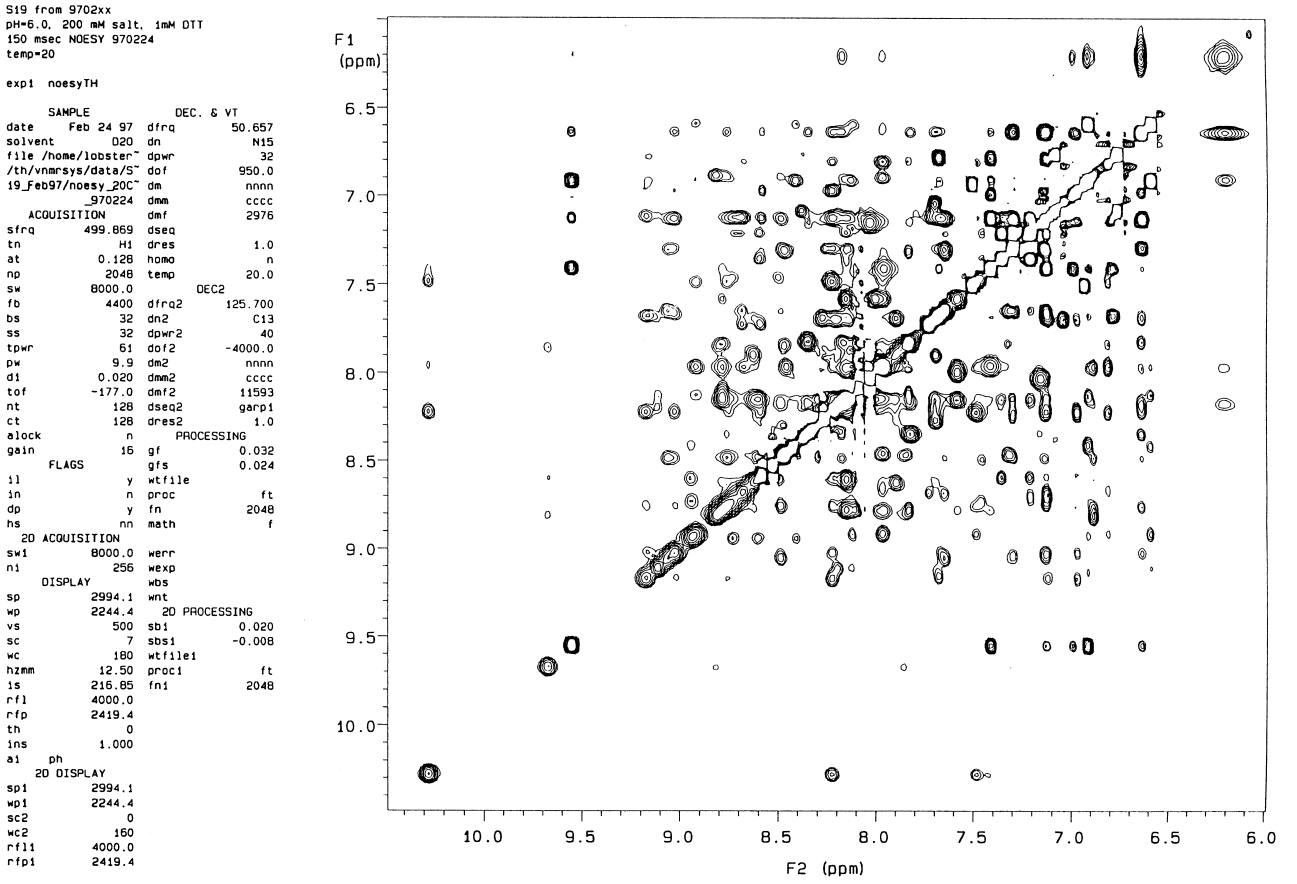

Fig. 3. Downfield region of a two-dimensional NOESY spectrum of the ribosomal protein S19.

tein pair S13–S19 in both *B. stearothermophilus* and *E. coli* [9].

The molecular mass of the purified recombinant protein determined by SDS-PAGE [19] electrophoresis in denaturing conditions (10 500) corresponds to that calculated from the amino acid sequence (10 569). Preparations of the protein purified by one-step chromatography on a CM-Sepharose column at pH 5.2 turned out to be unstable during storage at 4°C.

The analysis of the rpS19 preparation by SDS-PAGE after a month's storage in the refrigerator showed the presence of two bands, corresponding to 7500 and 3000. Perhaps the preparations contained a protease contamination invisible on PAG [19] electrophoregrams. Mild digestion with trypsin at 0°C also generated a fragment with molecular mass about 7500. The incubation time for the complete transition from the intact protein to the 7500 fragment was about 1 h. Trypsinolysis at 37°C produced a fragment of the same molecular mass but in a shorter time.

The N-terminus of the large stable protein fragment (S19fr) was determined, and S19fr was overproduced in *E. coli*. The high quality of preliminary NMR spectra of purified S19fr encouraged us to look for conditions for additional purification of the protein. Addition of a chromatographic step at different pHs to the purification procedure made it possible to avoid protease contamination. Finally, the purified protein could be kept at room temperature for more than a week without degradation. This protein preparation was used for NMR studies.

NMR spectra of purified and concentrated S19 (Fig. 2) showed that concentration and purity of this preparation appeared to be sufficient for a complete structure determination. The linewidths and resonance dispersion were characteristic of a folded protein with an approximate molecular mass of 10 000. The spectrum also suggests a high content of α-helical secondary structure. The dispersion in the amide proton region (around 8.4 ppm) was rather narrow and the majority of these resonances had a chemical shift of less than 9.2 ppm. Moreover, only three or four resonances were observed in the chemical shift region of 4.8–5.5 ppm, whereas this region normally is substantially more crowded with proteins with a high content of β-sheet secondary structure.

The downfield region of the two-dimensional NOESY spectrum is shown in Fig. 3. A large number of resolved NOEs can be observed sustaining the possibility of a future structure determination. In addition, the large number of NOEs between amide proton resonances and the fact that many resonances show two equally strong NOE connectivities (e.g. the resonance at 9.5 ppm) support the conclusion that S19 is a predominantly helical protein.

Acknowledgements: We thank Dmitryi Shcherbakov for the restriction map of pBR322-tthS8 and Katarina Wallgren for technical assistance. This work was supported by the Russian Foundation for Fundamental Research (94-04-12046-a), the Russian Academy of Sciences, the Swedish Natural Science Research Council (NFR), the Swedish Royal Academy of Sciences (KVA), NUTEK, the International Science Foundation (MTR000) and Cyanamid Company (USA). The research of M.G. was supported in part by an Interna-

tional Research Scholar's award from the Howard Hughes Medical Institute.

References

[1] Appelt, K., Dijk, J., Reinhardt, R., Sanhuesa, S., White, S.W., Wilson, K.S. and Yonath, A. (1981) J. Biol. Chem. 256, 11787–11790.
[2] Ramakrishnan, V. and Gerchman, S.E. (1991) J. Biol. Chem. 266, 880–885.
[3] Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) Methods Enzymol. 185, 61–89.
[4] Garber, M.B., Agalarov, S.Ch., Eliseikina, I.A., Sedelnikova, S.E., Tishchenko, S.V., Shirokov, V.A., Yusupov, M.M., Reshetnikova, L.S., Trakhanov, S.D., Tukalo, M.A. and Yaremchuk, A.D. (1991) J. Crystal Growth 110, 228–236.
[5] Powers, T., Changchien, L.-M., Craven, G.R. and Noller, H.F. (1988) J. Mol. Biol. 200, 309–319.
[6] Buck, M.A. and Cooperman, B.S. (1990) Biochemistry 29, 5374–5379.
[7] Powers, T. and Noller, H.F. (1995) RNA 1, 194–209.
[8] Capel, M.S., Engelman, D.M., Freeborn, B.R., Kjeldgaard, M., Langer, J.A., Ramakrishnan, V., Schindler, D.G., Schneider, D.K., Schoenborn, B.P., Sillers, I.-Y., Yabuki, S. and Moore, P.B. (1987) Science 238, 1403–1406.
[9] Kruft, V., Bischof, O., Bergmann, U., Herfurth, E., Wittmann-Liebold, B. (1993) The Translational Apparatus. Structure, Function, Regulation, Evaluation, pp. 509–520.
[10] Zheltonosova, J., Melnikova, E., Garber, M., Reinbolt, J., Kern, D., Ehresmann, C. and Ehresmann, B. (1994) Biochimie 76, 71–77.
[11] Hanahan, D. (1983) J. Mol. Biol. 166, 557–580.
[12] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[13] Marmur, J. (1961) J. Mol. Biol. 3, 208–218.
[14] Kunkel, T.A., Roberts, J.D. and Zakour, R.A. (1988) Methods Enzymol. 154, 367–382.
[15] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[16] Vysotskaya, V., Tischenko, S., Garber, M., Kern, D., Mougel, M., Ehresmann, C. and Ehresmann, B. (1994) Eur. J. Biochem. 223, 437–445.
[17] Davydova, N.L., Gryaznova, O.I., Mashchenko, O.V., Vysotskaya, V.S., Jonsson, B.-H., Al-Karadaghi, S., Liljas, A. and Garber, M.B. (1995) FEBS Lett. 369, 229–232.
[18] Clackson, T. and Winter, G. (1989) Nucleic Acids Res. 17, 10163–10170.
[19] Laemmli, U.K. (1970) Nature 227, 680–685.
[20] Macura, A. and Ernst, R.R. (1980) Mol. Phys. 41, 95–117.
[21] Kromer, W.J., Hatakeyama, T. and Kimura, M. (1990) Biol. Chem. Hoppe-Seyler 371, 631–636.
[22] Wittmann-Liebold, B. and Greuer, B. (1980) FEBS Lett. 121, 105–112.
[23] Sanangelantoni, A.M., Bocchetta, M., Cammarano, P. and Tiboni, O. (1994) J. Bacteriol. 7703–7710.
[24] Nagano, Y., Ishikawa, H., Matsuno, R. and Sasaki, Y. (1991) Plant Mol. Biol. 17, 541–545.
[25] McLaughlin, W.E. and Larrinua, I.M. (1987) Nucleic Acids Res. 15, 3932.
[26] Conklin, P.L. and Hanson, M.R. (1991) Nucleic Acids Res. 19, 2701–2705.
[27] Klein, H., Gervais, C. and Suh, M. (1990) Nucleic Acids Res. 18, 3997.
[28] Amaldi, F., Beccari, E., Bozzoni, I., Luo, Z.X. and Pierandrei-Amaldi, P. (1982) Gene 17, 311–316.
[29] Baumgartner, S.W., Martin, D. and Chiquet-Ehrismann, R. (1993) Nucleic Acids Res. 21, 3897.
[30] Suzuki, K., Olvera, J. and Wool, I.G. (1990) Biochimie 72, 299–302.
[31] Etter, A., Aboutanos, M., Tobler, H. and Möller, F. (1991) Proc. Natl. Acad. Sci. USA 88, 1593–1596.