

A subunit of an archaeal DNA-dependent RNA polymerase contains the S1 motif

Doris Langer, Friedrich Lottspeich and Wolfram Zillig
Max-Planck-Institut für Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, Germany

Received November 17, 1993; Revised and Accepted January 14, 1994 EMBL accession no. X75411

We cloned and sequenced the gene for the fifth largest subunit, E, of the RNA polymerase (RNAP) of the extremely thermophilic archaeon *Sulfolobus acidocaldarius*. Here we demonstrate that its central third contains the 'S1 motif', a putative RNA-binding motif which is present in the initiation factors IF-1 and eIF-2α, in the eucaryal RNA helicase-like protein PRP22, in the *E.coli* polynucleotide phosphorylase (PNP) and, in four copies, in the ribosomal S1 proteins.

The three largest of the 13 subunits of *S.acidocaldarius* RNAP, A', A'' and B show strong sequence similarity to the two largest components of all three eucaryal RNAPs and, less so, of bacterial RNAPs (1), but the smaller components H (2), K, L and N (3) are homologous to components only found in eucaryal and not in bacterial RNAPs. Three of the *S.acidocaldarius* components sequenced so far, E, F and G seem to be unique to archaeal RNAPs.

The N-terminal sequence of subunit E isolated by SDS−PAGE followed by electroblotting was determined in a gas-phase sequencer according to Eckerskorn *et al.* (4). A degenerated oligonucleotide derived from that sequence was used as a ^{32}P-labeled probe to clone and identify the gene encoding subunit E. The gene is 774 nucleotides long and encodes a protein of 258 amino acids and 27 KD, which corresponds to the apparent molecular mass estimated from the mobility in SDS−PAGE. By screening protein sequence databases [FASTA program (5)] with this amino acid sequence we found significant similarity to the S1-motif, which is present in bacterial ribosomal protein S1 as a repeated sequence motif (6), in single copy in the helicase-like protein PRP22 of yeast (7), in procaryotic and eucaryotic initiation factors IF-1 and eIF-2α and in polynucleotide phosphorylase of *E.coli* (8). Figure 1 shows a schematic drawing of the location of the S1 motif within the members of the S1-family (7, altered). The similarities according to Dayhoff *et al.* (9) and the identity values for the S1 motif of subunit E and the S1 motifs of the other proteins are given for one example of each of the proteins. The central third of subunit E, starting with amino acid 85 showed similarity to the S1-motif. No sequences corresponding to the N-terminal third were found, but the last third of subunit E shows a presumable metal ion binding motif with the structure Cys-X_2-Cys-X_{10}-Cys-X_2-Cys.

The presence of the S1 motif in an archaeal RNAP subunit remains not understood, as long as the function of subunit E in the transcription process is unknown. The bacterial ribosomal protein S1 is involved in binding the messenger RNA to the ribosome by unwinding the putative helical region near the 3'-terminus of the 16S rRNA (6). The RNA-helicase-like protein PRP22 is required for release of the mRNA product from the spliceosome (7). A similar process might occur during transcription termination where the nascent RNA has to be released from the coding strand of the DNA. Subunit E might be involved in this process. The ability to bind RNA and resolve interaction between strands in helices might be mediated by the S1 motif and this module could therefore in the course of evolution have been introduced into functionally quite different proteins endowed with this capacity.

REFERENCES

1. Pühler,G., Lottspeich,F. and Zillig,W. (1989) *Nucleic Acids Res.* **7**, 4517−4534.
2. Klenk,H.-P., Palm,P., Lottspeich,F. and Zillig,W. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 407−410.
3. Langer,D., Hain,H., Thuriaux,P. and Zillig,W. (1993) in preparation.
4. Eckerskorn,C. *et al.* (1988) *Eur. J. Biochem.* **176**, 509−517.
5. Pearson,R. and Lipman,D.J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2444−2448.
6. Schnier,J. and Isono,K. (1982) *Nucleic Acids Res.* **10**, 1857−1865.
7. Company,M., Arenas,J. and Abelson,J. (1991) *Nature* **349**, 487−493.
8. Gribskov,M. (1992) *Gene* **119**, 107−111.
9. Dayhoff,M.O. *et al.* (1978) In Dayhoff,M.O. (ed.), *Atlas of Protein Sequence and Structure.* Natl. Biomed. Res. Found., Washington, **Vol. V, Suppl. 3**, pp. 345−352.
10. Schnier,J., Thamm,S., Lurz,R., Hussain,A., Faist,G. and Dobrinski,B. (1988) *Nucleic Acids Res.* **16**, 3075−3089.

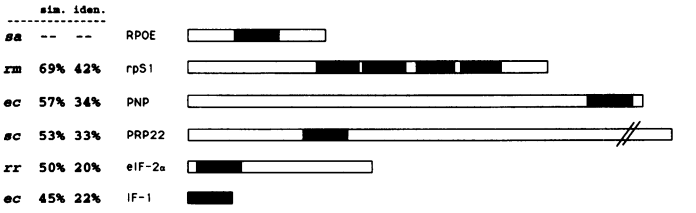

Figure 1. Schematic representation of proteins containing the S1 motif (shown as a black box) (7, altered). The order represents decreasing similarity to the S1 motif of the RNAP subunit E from *S.acidocaldarius* (RPOE). The values for identity and similarity given for one example of each group of proteins. rpS1 *rm*: ribosomal protein S1 from *Rhizobium melilotii*, (10), PNP *ec*: polynucleotide phosphorylase from *E.coli* (8), PRP22 *sc*: RNA-helicase-like protein from *Saccharomyces cereviseae* (7), eIF2α *rr*: initiation factor from rat (8), IF-1 *ec*: initiation factor from *E.coli* (8).