On a sequence similarity between ribosomal protein S5 and DNA binding protein II

Keith S. Wilson*, Makoto Kimura and Jan Dijk

*Department of Physics, University of York, York YO1 5DD, England and Max-Planck-Institut für Molekulare Genetik, Abt. Wittmann, Ihnestrasse 63–73, 1000 Berlin 33, Germany

Received 14 January 1985

Sequence similarity between ribosomal protein S5 and DNA binding protein II from eubacteria is reported and its significance discussed.

Ribosomal protein *DNA binding protein* *Sequence similarity*

1. INTRODUCTION

The amino acid sequence of ribosomal protein S5 from the 30 S subunit has been determined for both *Escherichia coli* [1] and *Bacillus stearothermophilus* [2]. The sequences of DNA binding protein II from a diverse set of eubacteria, *E. coli* [3], *B. stearothermophilus* [4], *Clostridium pasteurianum* [5], and the archebacterium *Thermoplasma acidophilum* [6], are also known. There is a high degree of homology within each of these families of proteins.

We here report a significant structural similarity between these two families, one family being an important part of the ribosomal biosynthetic machinery, and the other a major DNA binding protein of the prokaryotic cell. The similarity is restricted to a region of highly conserved residues in positions 39–48 of the *E. coli* DNA binding protein II. The sequence in this region is: -Gly-Asp-X-X-X-X-X-Gly-Phe-Gly-. In all 7 proteins this characteristic sequence is found within one or two residues to be the same distance from the N-terminus. While the two families of proteins have considerably different lengths (90 amino acids for DNA binding protein II and 166 for S5) we propose that there is at least a common structural feature in the conserved region above. The full similarity may extend through a much longer part of the structure.

Ribosomal protein S5 is an important component of the prokaryotic ribosome. It occurs at the interface between the two ribosomal subunits (30 S and 50 S). When the subunits are separated from one another in vitro S5 is found to be distributed between the large and the small subunits, both for *E. coli* and *B. Stearothermophilus* [7,8]. S5 is hence presumably involved in the interaction between the subunits. Its position close to the surface of the subunits leads to its easy extraction with low salt during our protein preparation under mild conditions [8]. S5 is important in the binding of a number of antibiotics to the ribosome, including spectinomycin, streptomycin, erythromycin and neamine. These results have been summarised [1].

DNA binding protein II occurs throughout the prokaryotic kingdom at a level of up to 100000 copies per cell [9]. There are two distinct molecular types present in *E. coli* [3], but only one in the other bacteria studied so far. We have presented the information known about DNA binding protein II in more detail elsewhere [10]. In brief, DNA binding protein II binds to double-stranded DNA, and causes the length of the DNA to decrease with the concomitant production of bead-like structures somewhat reminiscent of the nucleosomes of eukaryotic cells [11]. However, DNA binding protein II can also bind to single-stranded DNA and to RNA [12].

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/85/$3.30 © 1985 Federation of European Biochemical Societies

2. MATERIALS AND METHODS

The sequences were first aligned by hand using the single-letter amino acid code on strips of paper. The comparison was subsequently quantified using the ALIGN programme developed by Dayhoff and co-workers [13].

3. RESULTS AND DISCUSSION

The sequence of the 5 DNA binding proteins and of the two S5 proteins are presented in fig.1. The homology within each of the two families can be clearly seen. There are 55% identical residues between the S5 proteins from the two bacteria and the homology within the DNA binding proteins ranges from 32% to 83% identical residues.

Between the two families any similarity in amino acid sequence is much less apparent. We have noted however a common structural feature between residues 39 and 48 (numbered as the *E. coli* proteins), namely, the sequence -Gly-Asp-X-X-X-X-X-X-Gly-Phe-Gly-. This is totally conserved in all 7 sequences in the figure, with the exception that Asp is replaced by Gln in the *T. acidophilum* DNA binding protein and by Glu in *C. pasteurianum*.

Throughout the remainder of the sequence there is no comparable similarity between the two families as aligned in the figure. However the introduction of deletions and insertions in the sequences gives the possibility of aligning, for example, the conserved Leu residues at positions 6 in the DNA binding proteins and 9 in the S5 sequences and the Ala residues at positions 11 and 16, respectively. Similarly, the sequence -Ile-Glu-Val-Pro-

```
DNA binding protein II
                    5        10        15        20        25        30        35        40        45
E.c.(1)    M N K S Q L I D K I A A G A D I S K A A A G R A L D A I I A S V T E S L K E G D D V A L V
E.c.(2)    M N K T Q L I D V I A E K A E L S K T Q A K A A L E S T L A A I T E S L K E G D A V Q L V
B.st.      M N K T E L I N A V A E T S G L S K K D A T K A V D A V F D S I T E A L R K G D K V Q L I
C.p.       M N K A E L I T S M A E K S K L T K K D A E L A L K A L I E S V E E A L E K G E K V Q L V
T.a      M V G I S E L S K E V A K K A N T T Q K V A R T V I K S F L D E I V S E A N G G Q K I N L A
Ribosomal protein S5
E.c        A H I E K Q A G E L Q E K L I A V N R V S K T V K G G R I F S F T A L T V V G D G N G R V
B.st     M R R I N P N K L E L E E R V V A V N R V A K V V K G G R R L R F S A L V V V G D K N G H V
DNA binding protein II
                    50        55        60        65        70        75        80        85        90
E.c.(1)    G F G T F A V K E R A A R T G R N P Q T G K E I T I A A A K V P S F R A G K A L K D A V N
E.c.(2)    G F G T F K V N H R A E R T G R N P Q T G K E I K I A A A N V P A F V S G K A L K D A V K
B.st.      G F G N F E V R E R A A R K G R N P Q T G E E M E I P A S K V P A F K P G K A L K D A V K
C.p.       G F G T F E T R E R A A R E G R N P R T K E V I N I P A T T V P V F K A G K E F K D K V N K
T.a.       G F G I F E R R T Q G P R K A R N P Q T K K V I E V P S K K K F V F R A S S K I K Y Q Q
Ribosomal protein S5.
E.c.       G F G Y G K A R E V P A A I Q K A M E K A R R N M I N V A L N N G T L Q H P V K G V H T G ....
B.st.      G F G T G K A Q E V P E A I R K A I E D A K K N L I E V P I V G T T I P H E V I G H F G A ....
```

Fig.1. The amino acid sequences for DNA binding protein II and ribosomal protein S5 optimally aligned as described in the text. The conventional single-letter code is used. The highly conserved feature common to the sequences is indicated by the boxes. Only the N-terminal 90 residues are presented for S5. E.c., *Escherichia coli;* B.st., *Bacillus stearothermophilus;* T.a., *Thermoplasma acidophilum;* C.p., *Clostridium pasteurianum.*

Table 1

The alignment scores in units of standard deviation as defined in [13] for the comparison of the S5 with the DNA binding protein II sequences. There are two homologous proteins in *E. coli* labelled (1) and (2) respectively.

	DNA binding protein II				
	Eubacteria				Archebacterium *T. acidophilum*
	E. coli (1)	*E. coli* (2)	*B. stearothermophilus*	*C. pasteurianum*	
E. coli S5	1.18	1.23	0.97	−0.84	1.01
B. stearothermophilus S5	3.56	5.21	2.93	0.86	−0.84

present in positions 69–72 in *T. acidophilum* is identical to that present in positions 71–74 in *B. stearothermophilus* S5. We have not introduced these rather arbitrary deletions or insertions in the figure as there is no totally satisfactory algorithm for objectively weighting such modifications against the improved alignment produced.

We have tried to quantify the similarity using the ALIGN programme of Dayhoff and co-workers [13]. The sequences of DNA binding proteins were aligned with the first 100 residues of the two S5 proteins, using the genetic code matrix with a break penalty of 50. The results are summarised in table 1. All but two calculations *(E. coli* and *B. stearothermophilus* S5 proteins with the DNA binding protein of the archebacterium *T. acidophilum)* give the sequence alignments shown in fig.1. Alignment scores, as defined in [13], greater than 1 standard deviation were obtained for 5 out of the 8 pairs of eubacterial DNA binding and S5 proteins. This supplies weak, but significant, support of our hypothesis that the two groups of proteins are homologous.

The archebacterial DNA binding protein sequences while being about 30% identical to their eubacterial conterparts were not aligned with the S5 proteins as in fig.1 by the programme. ALIGN thus provides no evidence for significant homology for this protein with S5. Given the extensive period of divergence of eubacteria and archebacteria, and of the S5 and DNA binding proteins from their proposed ancestor, this is perhaps not a surprising result.

We conclude that the similarity in amino acid sequences in the conserved region 39–48 will reflect a three-dimensional structural similarity in this region of both families. Moreover, the occurrence of this region at close to identical distances from the N-terminus in the two sets, together with the possibility of increased similarity over more extended lengths of chain if deletions or insertions are allowed, leads to the possibility of a more extensive structural relation between these two families.

Any homology in sequence over the 90 N-terminal amino acids is reflected in a very low number of identical residues for the alignment shown in fig.1. We are currently carrying out X-ray crystallographic studies of both S5 and the DNA binding protein II from *B. stearothermophilus:* the three-dimensional structure of the former has been analysed at 3 Å resolution [10] and the latter at 6 Å [14]. The conformation of the conserved sequence is a strand of antiparallel β-pleated sheet, with the two glycines occupying key positions in sharp β-bends at either end of the strand. The conformation in S5 awaits analysis of the structure at higher resolution. The hypothesis of homology will be stringently tested by the comparison of the tertiary structure of these proteins.

A structural relation between a DNA binding and a ribosomal protein would be an interesting result for the evolution of these proteins. It is of note in this regard that S5 has been reported to bind to ribosomal RNA [15,16] and that DNA binding protein II can also bind to RNA [12] and to both 70 S ribosomes and to 30 S and 50 S subunits [10,17]. In our purification system we find some DNA binding protein II attached to the ribosome after digesting away the cellular DNA, perhaps bound to ribosomal RNA in a manner comparable to S5.

REFERENCES

[1] Wittmann-Liebold, B. and Greuer, B. (1978) FEBS Lett. 95, 91-98.
[2] Kimura, M. (1984) J. Biol. Chem. 259, 1051-1055.
[3] Mende, L., Timm, B. and Subramanian, A.R. (1978) FEBS Lett. 96, 395-398.
[4] Kimura, M. and Wilson, K.S. (1983) J. Biol. Chem. 258, 4007-4011.
[5] Kimura, M., Kimura, J. and Zierer, R. (1984) FEBS Lett. 175, 208-212.
[6] Delange, R.J., Williams, L.C. and Searcy, D.G. (1981) J. Biol. Chem. 256, 905-911.
[7] Sommer, A. and Traut, R.R. (1976) J. Mol. Biol. 106, 995-1015.
[8] Appelt, K. and Dijk, J. (1985) J. Biol. Chem., in press.
[9] Geider, K. and Hoffmann-Berling, H. (1981) Annu. Rev. Biochem. 50, 233-260.
[10] Tanaka, I., Appelt, K., Dijk, J., White, S.W. and Wilson, K.S. (1984) Nature 310, 376-381.
[11] Rouviere-Yaniv, J., Yaniv, M. and Germond, J.E. (1979) Cell 17, 255-274.
[12] Berthold, V. and Geider, K. (1976) Eur. J. Biochem. 71, 443-449.
[13] Dayhoff, M.O., Schwartz, R.M. and Orcutt, B.C. (1979) in: Atlas of Protein Sequence and Structure, vol.5, suppl. 3 (Dayhoff, M.O. ed.) pp. 345-363, National Biomedical Research Foundation, Washington, DC.
[14] White, S.W., Appelt, K., Dijk, J. and Wilson, K.S. (1983) FEBS Lett. 163, 73-75.
[15] Hochkeppel, H.-K., Spicer, E. and Graven, G.R. (1976) J. Mol. Biol. 101, 155-170.
[16] Littlechild, J., Dijk, J. and Garrett, R.A. (1977) FEBS Lett. 74, 292-294.
[17] Suryanarayana, T. and Subramanian, A.R. (1978) Biochim. Biophys. Acta 520, 342-357.