The primary structure of the DNA-binding protein II from *Clostridium pasteurianum*

Makoto Kimura, Junko Kimura and Reiner Zierer

Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, Ihnestrasse 63–73, D-1000 Berlin 33 (Dahlem), Germany

Received 2 August 1984

The complete amino acid sequence of the *Clostridium pasteurianum* DNA-binding protein II (DNAb-II) has been determined. The molecule contains 91 amino acid residues and has an M_r of 10133. Sequence data were obtained from manual Edman degradation, using the DABITC/PITC double-coupling of the tryptic, peptic, chymotryptic and *Staphylococcus* protease peptides. A comparison of the amino acid sequence of the *C. pasteurianum* DNAb-II with those of the DNAb-II from *Escherichia coli, Bacillus stearothermophilus, Thermoplasma acidophilum* and *Pseudomonas aeruginosa* shows that the *C. pasteurianum* protein is more homologous to that of *B. stearothermophilus* (60%) than to that of *E. coli* (45%). All DNAb-II proteins have identical sequences Gly-Phe-Gly-X-Phe at positions 46–50 and Arg-Asn-Pro-X-Thr at positions 61–65.

Amino acid sequence *DNA-binding protein II* *Protein evolution* Clostridium pasteurianum

1. INTRODUCTION

DNA-binding protein II (DNAb-II), previously called NS1 and NS2 which are homologous proteins, were isolated from *E. coli* [1] and have been studied as suitable molecules to probe the interaction of protein with DNA.

As reviewed in [2], DNAb-II exists in $\sim 10^5$ copies per *E. coli* cell and consists of 90 amino acid residues, as revealed by the amino acid sequence [3]. In solution the protein occurs in an oligomeric form. As to the functional property of these proteins, it has been found that they can stabilize DNA by causing an increase in both transition temperature and extent of renaturation [4], and that they inhibit the transcription of several DNAs [5].

Several DNA binding proteins have been isolated from different organisms [6–8]. One of these proteins, DNAb-II from *Bacillus stearothermophilus*, has been crystallized [6], and authors in [9] have recently constructed an X-ray crystallographic model from an electron-density map of the molecule at 3 Å resolution.

Simultaneously, the amino acid sequence of the *B. stearothermophilus* DNAb-II was determined [10] and compared with that of NS1/NS2 and other homologous proteins. The sequences show that there is considerable homology among these proteins in spite of the evolutionary distance of the organisms. Therefore, it appears that the DNAb-II proteins are excellent molecules to study the protein/nucleic acid interaction and also to examine the evolution of these proteins. Hence, to collect the sequence data and also to refine the structure model of the *B. stearothermophilus* DNAb-II by detailed comparison, we have extended the amino acid sequence work to DNAb-II from *Clostridium pasteurianum*, a gram-positive eubacterium. Here, we present the primary structure of this protein and compare its primary structure with those of other organisms.

2. MATERIALS AND METHODS

The DNA-binding protein II was isolated from *C. pasteurianum* (strain NCIB 9486, obtained from CAMR, Porton, England) using a modification of the previous method [6]. More details of the purification procedure will be given in [11].

Published by Elsevier Science Publishers B.V.

Enzymes and other materials for sequencing were used as described in [12]. The methods of sequence analysis were the same as in [10].

3. RESULTS AND DISCUSSION

3.1. *Sequence determination*

As previously described [10] the DNA binding protein II from *B. stearothermophilus* was successfully sequenced by analysis of the tryptic and peptic peptides. However, in the case of the protein derived from *C. pasteurianum* this method was not sufficient to complete the sequence analysis. Difficulties were encountered in the region, positions 67–80, since the tryptic peptide T7 (pos.42–53) comigrated with T13 (pos.67–80) on thin-layer fingerprints. Furthermore, the peptic peptide P9 (pos.68–79) was difficult to isolate. Therefore, only tentative sequences of positions 67–80 were available, while the amino acid sequence of positions 42–66 could be determined by sequencing the peptic peptides P5 to P8.

The sequence of the region (pos.67–80) was finally established in peptides derived from chymotryptic digestion of DNAb-II. This digest produced 8 peptides, and the resulting peptides could be purified by fingerprinting. Manual solid-phase sequencing of the peptides C5 (pos.51–79) and C5a (pos.65–79) completed the sequence of the remaining region, as given in fig.1. To confirm

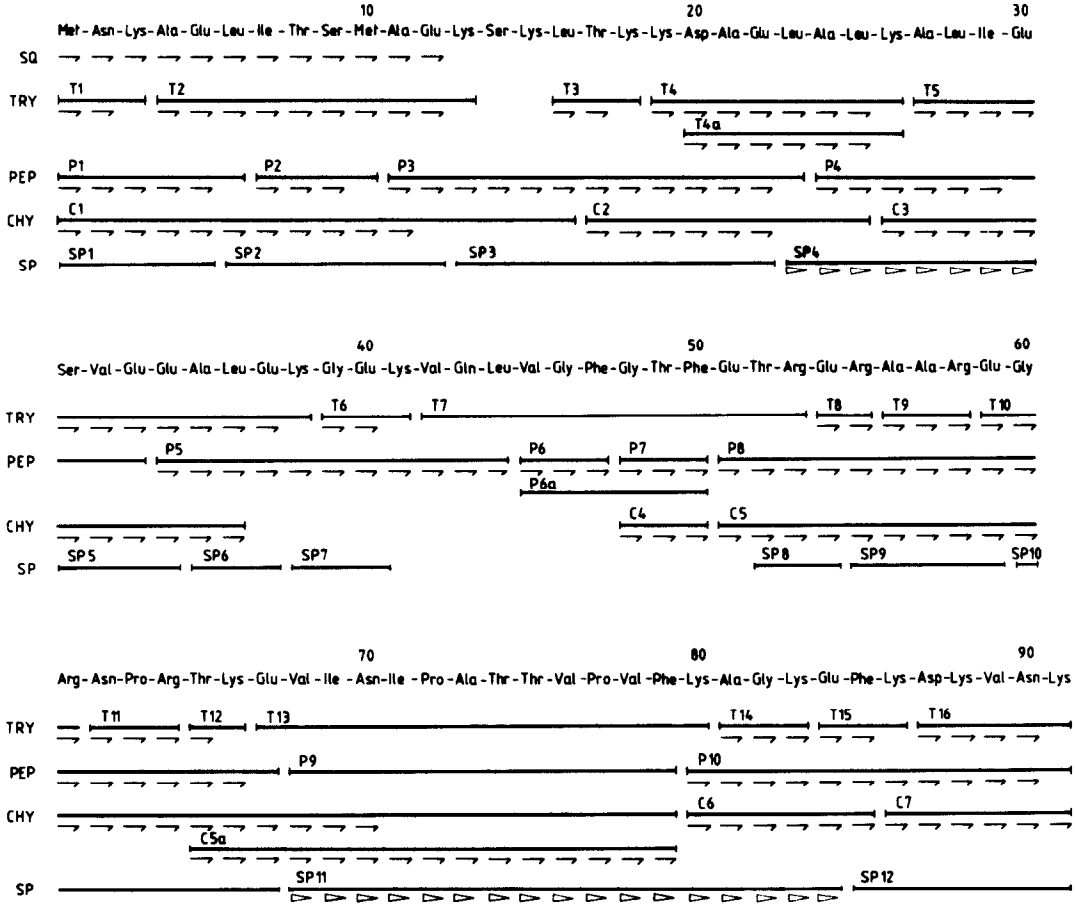

Fig.1. The amino acid sequence of the DNA binding protein II from *C. pasteurianum*. Sequence data on individual peptides are indicated as follows: —, sequenced by DABITC/PITC double coupling method; ▻, sequenced by manual solid-phase procedure. SQ indicates degradation of intact protein. TRY, PEP, CHY and SP indicate peptides derived from digestion with trypsin, pepsin, chymotrypsin and *S. aureus* protease, respectively.

the amino acid sequence unambiguously, the protein was digested with *Staphylococcus* protease. The resulting peptides (SP1 and SP12) were separated by fingerprinting and sequenced. The results obtained were in precise agreement with the sequence presented in fig.1.

3.2. *Characterization of the protein*

The complete primary structure of the DNA-binding protein II from *C. pasteurianum* is shown in fig.1. The protein contains 91 amino acid residues with the amino acid composition: Asp_2 Asn_4 Thr_7 Ser_3 Glu_{13} Gln_1 Pro_3 Gly_5 Ala_{10} Val_7 Met_2 Ile_4 Leu_7 Phe_4 Lys_{14} Arg_5. From the sequence an M_r of 10133 is calculated. Lysine and glutamic acid are the most abundant amino acids in this protein. The interesting feature of the amino acid sequence of the protein, which seems to be shared by all DNAb-IIs, is that almost all charged residues occur consecutively at 3 regions in positions 12–41, 51–67, and 80–91, with an intervening cluster of hydrophobic strings which occur at positions 42–50 and 68–79.

3.3. *Comparison of 6 DNA-binding proteins*

Fig.2 shows a comparison of the sequence of *C. pasteurianum* DNAb-II with the DNA-binding proteins of *B. stearothermophilus, E. coli, Pseudomonas aeruginosa* and *Thermoplasma acidophilum*. As shown in fig.2, the *C. pasteurianum* sequence can easily be aligned with the other proteins. (Note that the *P. aeruginosa* protein has not been completely sequenced.)

It is especially interesting to compare the primary structure of the *C. pasteurianum* DNA-binding protein with that of *B. stearothermophilus* for which also the tertiary structure is known. It turns out that all amino acid residues which are in the *B. stearothermophilus* protein involved in an intra- or inter-molecular interaction, namely Ala-11, Ala-21, Ala-24, Leu-36, Gly-39, Leu-44, Phe-47, Phe-50 and Phe-79, are conserved in the

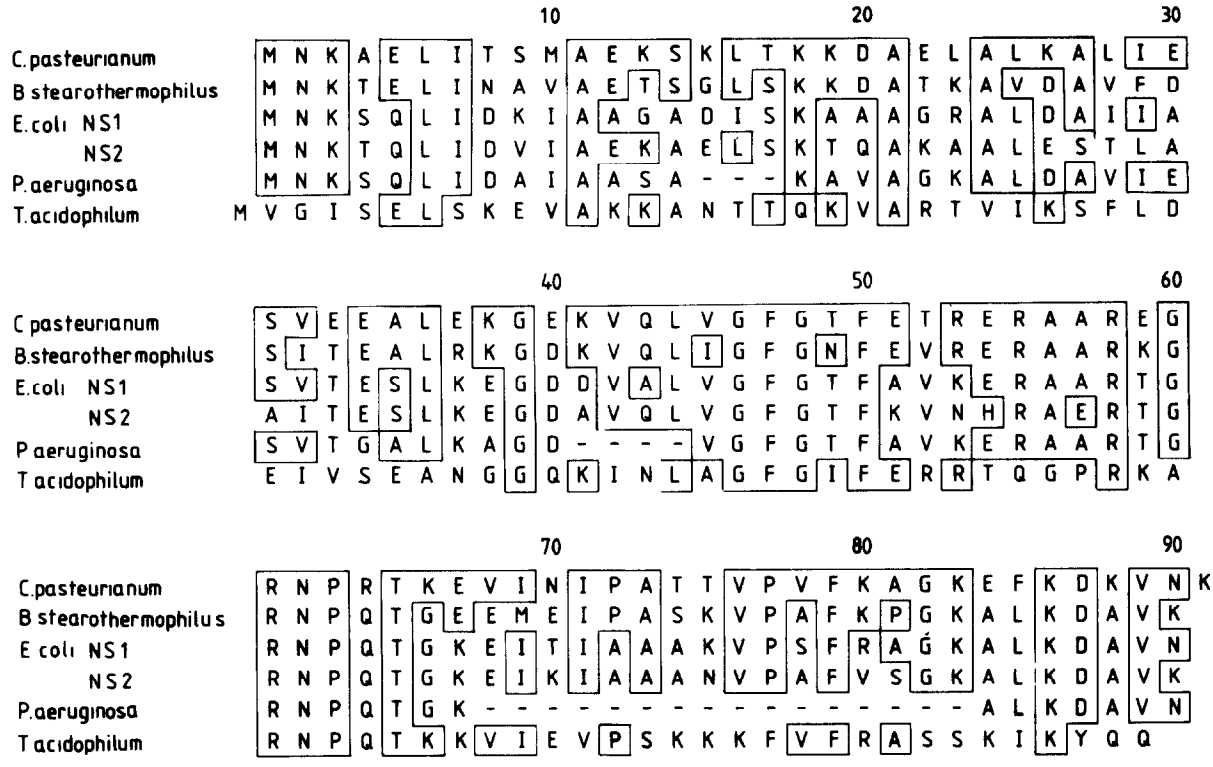

Fig.2. Comparison of the amino acid sequences of DNA binding protein II from prokaryotes. Proteins compared are from *C. pasteurianum, B. stearothermophilus* [10], *E. coli* NS1 and NS2 [3], *P. aeruginosa* [8] and *T. acidophilum* [14]. The residues identical to the *C. pasteurianum* protein are boxed.

Table 1

Degree of homology

	C. pasteurianum	*B. stearothermophilus*	*E. coli* NS1	*E. coli* NS2	*T. acidophilum*
C. pasteurianum	–				
B. stearothermophilus	60	–			
E. coli NS1	47	58	–		
E. coli NS2	45	59	69	–	
T. acidophilum	27	32	27	27	–

The percentage of identical amino acids at identical positions is given for each pair of the various DNA-binding proteins

C. pasteurianum protein, with the exception of Ile-32 which is substituted by Val-32. The DNA binding sites which have been deduced from the crystallographic model of the *B. stearothermophilus* protein at the regions in positions 53–63 and 75–91 are also conserved in the *C. pasteurianum* protein and contain clusters of charged amino acid residues.

On the other hand, the *C. pasteurianum* protein is characterized by an additional lysine residue at the C-terminus. Furthermore, at residues 26, 33, 37, 59, 64, 66, 84 and 88, where most other eubacterial proteins have the same amino acids, the *C. pasteurianum* protein has amino acids with difference in charge. The significance of these replacements will be better understood if the 3-dimensional structure of the *C. pasteurianum* protein is known, and crystals of this protein have already been obtained (R. Zierer, personal communication).

A matrix showing the degree of homology between the DNAb-II proteins from various organisms is presented in table 1. In this table the value for the *P. aeruginosa* protein is not included since the sequence data for this protein are incomplete. The protein from *C. pasteurianum* shows the best homology to the protein from *B. stearothermophilus* (60%), a slightly lower homology to the protein pairs from *E. coli*, NS1 (47%) and NS2 (45%), and much less homology to the protein from *T. acidophilum* (27%).

From this individual comparison it is likely that the DNAb-II sequences from the gram-positive eubacteria (*B. stearothermophilus* and *C. pasteurianum*) show a higher homology with each other than with the proteins from the gram-negative eubacteria (*E. coli*). A similar result derives from comparison of ribosomal L7/L12 proteins from different sources [13].

ACKNOWLEDGEMENTS

We would like to thank Dr H.G. Wittmann for his interest and encouragement, Drs B. Wittmann-Liebold, J. Dijk and K. Wilson for critically reading the manuscript and Mr D. Kamp for performing the amino acid analyses.

REFERENCES

[1] Suryanarayana, T. and Subramanian, A.R. (1978) Biochim. Biophys. Acta 520, 342–357.
[2] Geider, K. and Hoffmann-Berling, H. (1981) Ann. Rev. Biochem. 50, 233–260.
[3] Mende, L., Timm, B. and Subramanian, A.R. (1978) FEBS Lett. 96, 395–398.
[4] Miano, A., Losso, M.A., Gianfranceschi, G.L. and Gualerzi, C.O. (1982) Biochem. Internat. 5, 415–422.
[5] Losso, M.A., Miano, A., Gianfranceschi, G.L. and Gualerzi, C.O. (1982) Biochem. Internat. 5, 423–427.
[6] Dijk, J., White, S.W., Wilson, K.S. and Appelt, K. (1983) J. Biol. Chem. 258, 4003–4006.
[7] DeLange, R.J., Green, G.R. and Searcy, D.G. (1981) J. Biol. Chem. 256, 900–904.
[8] Hawkins, A.R. and Wootton, J.C. (1981) FEBS Lett. 130, 275–278.
[9] Tanaka, I., Appelt, K., Dijk, J., White, S.W. and Wilson, K.S. (1984) Nature, in press.
[10] Kimura, M. and Wilson, K. (1983) J. Biol. Chem. 258, 4007–4011.
[11] Zierer, R. (1984) Dissertation, Freie Universität Berlin, in preparation.

[12] Wittmann-Liebold, B. and Lehmann, A. (1980) in: Methods in Peptide and Protein Sequence Analysis (Birr, C. ed.) pp.49–72, Elsevier, Amsterdam, New York.

[13] Yaguchi, M., Matheson, A.T., Visentin, L.P. and Zuker, M. (1980) in: Genetics and Evolution of RNA Polymerase, tRNA and Ribosomes (Osawa, S. et al. eds) pp.585–599, University of Tokyo Press, Elsevier, Amsterdam, New York.

[14] DeLange, R.J., Williams, L.C. and Searcy, D.G. (1981) J. Biol. Chem. 256, 905–911.