Volume 273, number 1,2, 59–62 FEBS 09006 October 1990

The primary structure of DNA binding protein II from the extreme thermophilic bacterium *Thermus thermophilus**

Rainer Zierer and Dora Choli

Max Planck Institute for Molecular Genetics, Abt. H-G Wittmann, Berlin Dahlem, FRG

Received 20 August 1990

The primary structure of DNA binding protein II (DNA bp II) from the extreme thermophilic bacterium *Thermus thermophilus* has been established by combination of manual and automated techniques. The protein has 95 residues and a molecular mass of 11 843. Comparison of the primary structure with the known sequence data of DNA bp II from *Clostridium pasteurineum*, *Baccillus stearothermophilus*, *Escherichia coli*, *Rhizobium meliloti*, *Anabena*, *Thermoplasma acidophilum*, *Pseudomonas aeruginosa* and *Bacillus caldolyticus* reveals a clear homology among these small basic proteins. In particular, two short sequences in the middle and C-terminal part of the proteins (residues N-Gly-Phe-Gly-X-Phe and Pro-X-Thr at positions 46–51 and 63–65, respectively) are completely conserved.

Thermophilic bacteria; DNA-binding protein II; Primary structure; Homology

1. INTRODUCTION

A number of proteins appear to be involved in the organization of the prokaryotic genome, among them the DNA binding protein II (DNA bp II) [1]. This protein binds non-sequence specifically to the DNA and may carry out a role comparable to that of histones in eukaryotes. The basic protein was first isolated from *E. coli* [2]. Electron microscopy has shown that the protein is able to form bead-like structures with prokaryotic DNA [3] in a manner similar to the chromatin structures of eukaryotic DNA [4], suggesting a histone-like interaction with DNA.

DNA bp II generally have monomeric molecular masses of approximately 9500. The primary sequence has been determined for DNA bp II from a number of bacteria; for references see [5]. The three-dimensional structure of DNA bp II from *B. stearothermophilus* has been determined at 3 Å by X-ray crystallography and a model has been proposed for the interaction of the protein with DNA [4].

Isolation and crystallization of the *T. thermophilus* DNA bp II have been reported previously [6]. *T. thermophilus* can grow at a temperature of up to 80°C [7]. It is to be expected that the DNA bp II will be particularly well-ordered and stable, and thus ideal for three-dimensional crystallographic analysis of its structure. DNA bp II from *T. thermophilus* differs in several ways from the homologous proteins of the other organisms: upon nucleic acid binding, conformational changes in the protein's tertiary structure have been observed using proton magnetic resonance spectroscopy [6]. In contrast to the homologous proteins from *E. coli* [2] we found a destabilising effect of DNA bp II on DNA [6]. The complete amino acid sequence of *T. thermophilus* is presented and its homology discussed to other DNA bp II.

Correspondence address: R. Zierer, Birkenweg 7, 3413 Moringen, FRG

*This work represents for R.Z. partial fulfillment of the requirements for the Degree of Doctor of Philosophy at the Free University of Berlin, FRG

2. MATERIALS AND METHODS

All sequence reagents were of the highest purity available. *T. thermophilus* cells, strain HB8, where grown [6] and the purification of DNA bp II was carried out as described [6]. DNA bp II was digested with TPCK-trypsin, cleaved with CNBr [8] and the peptides were isolated as previously described [9]. The COOH-terminal fragment produced by CNBr was further digested with *Staphylococcus aureus* proteinase V8 [9]. Sequence determination of the intact protein and isolated peptides was carried out by manual Edman degradation employing the 4-*N*,*N*-dimethylaminoazobenzene-4′-isothiocyanate/phenylisothiocyanate double coupling procedure [10]. CNBr peptides were further submitted to pulse-liquid phase sequencing [8] on an Applied Biosystems, model 477A, equipped with a PTH-ad analyser, model 120. The secondary structure of DNA bp II was predicted based on the amino acid sequence using different methods [11].

3. RESULTS AND DISCUSSION

3.1. *Sequence determination*

In this subsection (Fig. 1), we number the sequence of *T. thermophilus* DNA bp II from 1 at the N-terminus. In later sections and figures we number the same protein from −5 in order to allow direct alignment of the

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

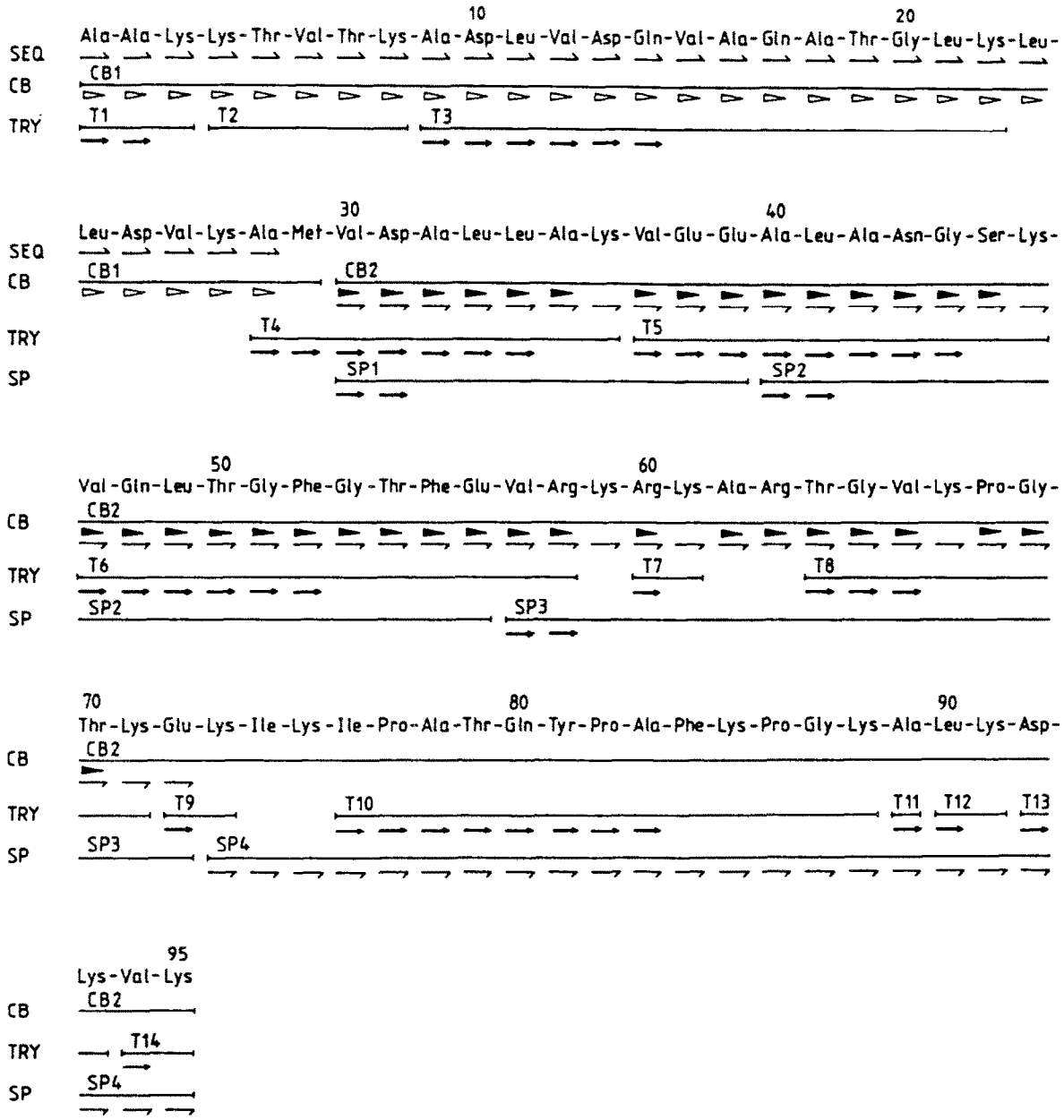

Fig. 1. Amino acid sequence of DNA bp II from *T. thermophilus*. The sequence of individual peptides and intact proteins is indicated as follows: (→) sequenced automatically using a pulse gas-liquid phase sequencer; (→) manual liquid-phase DABITC/PITC double coupling method; (▷) solid-phase sequencing after homoserine-lactone attachment to aminopropyl glass; (▶) solid-phase sequencing after attachment to diisothyocyanate-aminopropyll glass. SEQ indicates sequencing of the intact protein. TRY, CB and SP indicate peptides derived from cleavage with trypsin, CNBr or *Staphylococcus aureus* proteinase, respectively.

sequence with those of *B. stearothermophilus, E. coli*, etc. Analysis of tryptic digest revealed 13 peptides (T1–T13). The two CNBr fragments (positions 1–29 and 30–95) were separated by reversed phase HPLC. Both the N-terminal (CB1; 1–29) and the COOH-terminal peptides (CB2) were sequenced automatically by solid phase techniques or with a gas-liquid phase sequencer (Fig. 1). The C-terminal peptide was digested with *Staphylococcus aureus* proteinase. Four fragments (SP1–SP4) were separated using RP-HPLC. The complete amino acid sequence of *T. thermophilus* DNA bp II is shown in Fig. 1. The amino acid composition derived from the sequence is in good agreement with that obtained from the total hydrolysis of the protein (Table I).

3.2. *Sequence comparison*

Please note we now number the sequence of *T. thermophilus* DNA bp II from −5 at the N-terminus in Fig. 2 in order to align it with the previously determined structures of the *B. stearothermophilus* and *E. coli* DNA bp II. The high degree of conservation is immediately obvious. Using the *B. stearothermophilus*

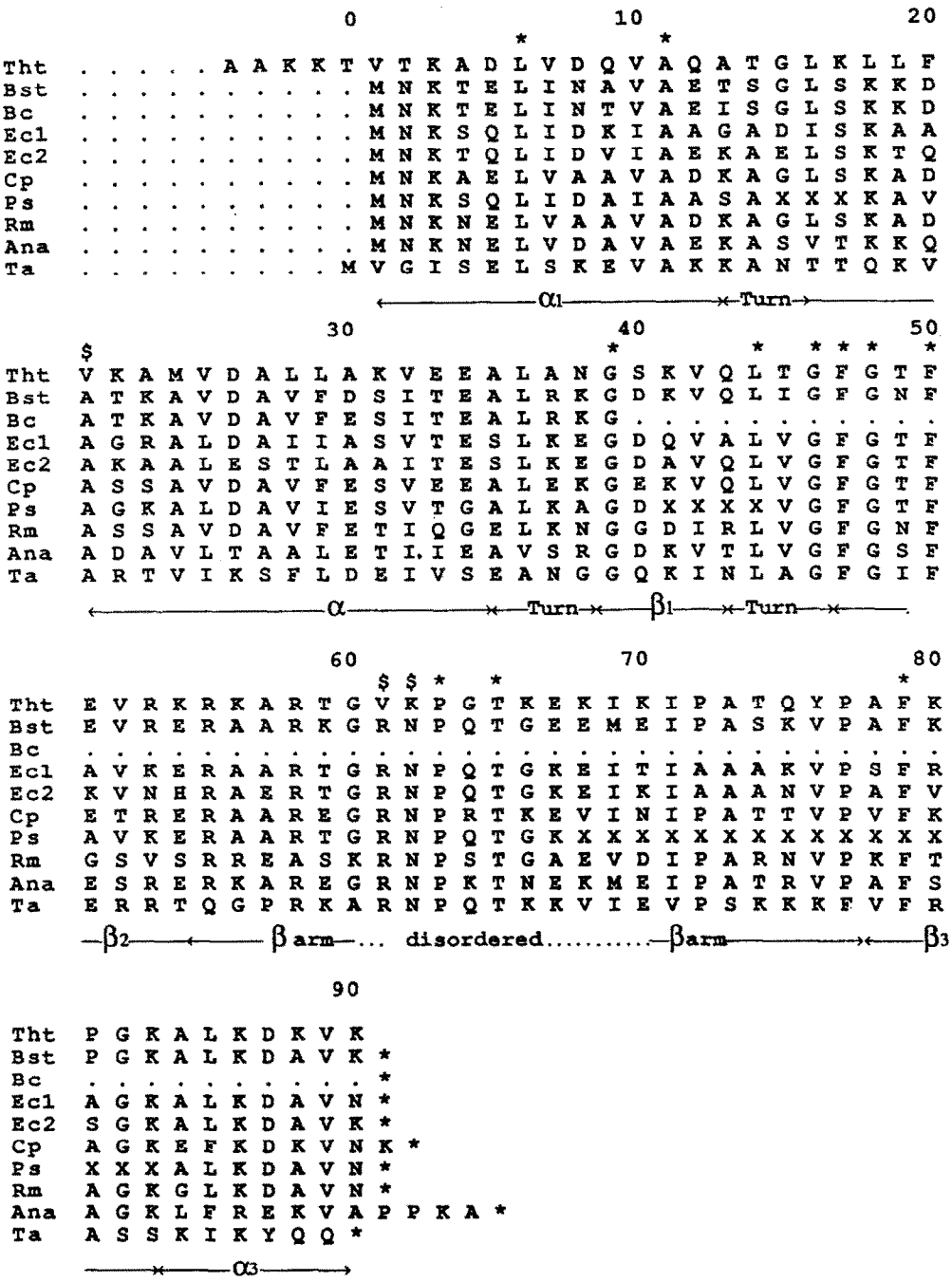

Fig. 2. Structural homology between DNA bp II from *T. thermophilus* and DNA bp II from other organisms, for references see [5]. The abbreviations in the figure are: Tht, *T. thermophilus*; Bst, *B. stearothermophilus*; Bc, *B. caldolyticus*; Ec1 and Ec2, *E. coli* (there are two copies of the protein present); Cp, *C. pasteurineum*; Ps, *P. aeruginosa*; Rm, *Rhizobium meliloti*; Ana, Anabena; Ta, *T. acidophilum*. The standard one letter code is used for the amino acids. The *Pseudomonas* sequence has been only partially determined and only the N-terminal sequence has been published for *B. caldolyticus*. The secondary structural elements from the known tertiary structure in *B. stearothermophilus* DNA bp II are shown. Totally conserved residues are indicated with a '*'. Residues conserved in all sequences except *T. thermophilus* are marked with a '$'.

protein as reference, the percentage of identical residues is 32%, 47%, 59% and 60% for the *Th. acidophilum, T. acidophilum, E. coli* and *C. pasteurineum* proteins, respectively.

3.3. *Structural discussion of DNA bp II (see Fig. 2)*

The three-dimensional structure of DNA bp II from *B. stearothermophilus* has been determined by X-ray crystallography. Therefore it is taken as reference molecule for further discussion because the highly conserved positions in the sequences can be understood from this known structure [4].

Comparison of the sequence of *T. thermophilus* DNA bp II with that of the other proteins shows the substitutions to be more in the N-terminal region of the molecule rather than in the central or COOH-terminal

Table I

Amino acid analysis of DNA bp II from *T. thermophilus*

Amino acid	Number of residues derived be sequence	Analysis
Asp	5	5.4
Asn	1	-
Glu	4	8.0
Gln	4	-
Ser	1	0.9
Gly	7	6.8
Thr	8	7.9
Arg	3	2.9
Ala	14	13.8
Tyr	1	0.8
Met	1	1.1
Val	10	5.1
Phe	3	2.4
Ile	2	1.2
Leu	9	7.1
Lys	18	15.5
Pro	4	n.d.

The values given are not corrected for destruction of amino acids or incomplete hydrolysis; n.d. indicates residues not determined by amino acid analysis.

part. Two alanine residues are highly conserved in the helices of the other proteins, namely Ala-11 and Ala-21. In *T. thermophilus* DNA bp II Ala-11 is conserved, but Ala-21 changed to valine.

The β-1 and β-2 sheets are linked by a short, highly conserved region, namely N-Gly-Phe-Gly-X-Phe at position 46–50, which is present in all DNA bp II so far sequenced.

Hydrophobic amino acids, also conserved, are found at Leu-6, Ala-11, Leu-16, Met-24, Leu-36, Leu-44, Val-52, Ile-71, Leu-85 and Val-89. These residues are responsible for both inter- and intra-molecular stabilization of the polypeptide chain [6].

The nucleic acid binding site of DNA bp II is proposed from the model building studies to be located in the antiparallel β-pleated sheet region, namely at residues 53–70. This region is characterized by positively charged residues, namely Arg-53, Lys-54, Arg-55, Lys-56, Arg-58, Lys-62, Lys-70, which can interact with phosphate backbone of nucleic acids.

There are 5 extra residues at the N-terminus of *T. thermophilus* DNA bp II. The significance of this feature is not clear at the present time.

Based on the *T. thermophilus* DNA bp II sequence determined above the secondary structural elements of DNA bp II as predicted by four different methods [11] revealed an α-helix content of 36%. The α-helices are located at the N-terminus and COOH-terminus of the molecule. This is in good agreement with the experimental data obtained by circular dichroism [6].

Acknowledgements: Dr K.S. Wilson is acknowledged for helpful discussion on the sequence, Ulrike for excellent technical assistance, Dr Maio for reading the manuscript and Dr Wittmann for encouragement. R.Z. would like to thank his parents and his Aunt Ursel for financial support.

REFERENCES

[1] Geider, K. and Hoffmann-Berling, H.A. (1981) Annu. Rev. Biochem. 50, 233–260.
[2] Suryanaraya, T. and Subramanian, A.R. (1978) Biochim. Biophys. Acta 520, 342–357.
[3] Berthold, V. and Gaider, K. (1976) Eur. J. Biochem. 71, 443–449.
[4] Tanaka, I., Appelt, K., Dijk, J., White, S. and Wilson, K. (1984) Nature 310, 376–381.
[5] Flamm, E.L. and Weisberg, R.A. (1985) J. Mol. Biol. 183, 117–128.
[6] Zierer, R., Graote, M., Dijk, J. and Wilson, K.S. (1986) FEBS Lett. 194, 235–241.
[7] Oshima, T. and Imahori, K. (1974) Int. J. Syst. Bacteriol. 24, 104–112.
[8] Ashman, K. and Bosserholt, A. (1985) in: Modern Methods in Protein Chemistry (Tschesche, H. ed.) vol. 2, pp. 155–171, W. de Gruyter Verlag, Berlin.
[9] Wittmann-Liebold, B., Bauer, D. and Dogin, J. (1977) in: Solid Phase Methods in Protein Sequence Analysis, (Previero, A. and Coletti-Rievero, A.M. eds) pp. 219–233, INSERM.
[10] Chang, J.Y., Brauer, D. and Wittmann-Liebold, B. (1978) FEBS Lett. 93, 205–214.
[11] Rawlings, N., Ashman, K. and Wittmann-Liebold, B. (1983) Int. J. Peptide Protein Res. 22, 515–524.