Vol. 184, No. 1, 1992 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
April 15, 1992 Pages 31-35

CHARACTERIZATION OF A cDNA CLONE FROM THE HAEMOPARASITE *BABESIA BOVIS* ENCODING A PROTEIN CONTAINING AN "HMG-BOX"[1]

Brian P. Dalrymple and Jennifer M. Peters

Commonwealth Scientific and Industrial Research Organisation, Division of Tropical Animal Production, Meiers Rd., Indooroopilly, Queensland 4068, Australia

Received February 18, 1992

SUMMARY: The complete nucleotide sequence of a *Babesia bovis* cDNA clone encoding a protein containing an HMG-Box has been determined. The predicted protein of 97 amino acids has a molecular weight of 11,116. It exhibits approximately 45% overall amino acid identity with the *Saccharomyces cerevisiae* non-histone protein 6A (NHP6A) and approximately 57% identity in the HMG-Box. The *B. bovis* protein has been designated NHP1. Like HNP6A, and unlike most other HMG1 homologues, NHP1 does not have a basic or an acidic carboxy-terminal domain. The amino acid sequence of HNP1 is much less similar to HMG1 homologues of another protozoan, *Tetrahymena thermophila*, than to the HMG1 homologues identified in *S. cerevisiae*, plants and vertebrates. This suggests that the *T. thermophila* proteins may not be true HMG1 homologues, or that they may be evolving at a much faster rate.

The HMG-Box is an approximately 75 amino acid domain containing a DNA binding site (1). HMG-Boxes have been identified in a number of DNA-binding proteins including; nuclear chromosomal high mobility group proteins, HMG1 (see 2) and HMG2 (3), related mitochondrial proteins (4,5), transcription factors (6-10) and *Schizosaccharomyces pombe* mating type proteins (11). The HMG1 and HMG2 proteins from higher eukaryotes contain two HMG-Boxes and have a long, predominantly acidic carboxy-terminal domain (2). However, maize (12), soy bean (13), *Saccharomyces cerevisae* (14) and *Tetrahymena thermophila* (15) HMG1 homologues have only one HMG-Box. Additionally, the *T. thermophila* HMGB and HMGC proteins have a basic, rather than an acidic, carboxy-terminal domain (15). The *S. cerevisiae* HNP6A and HNP6B proteins and the mitochondrial proteins, mtTF1 and ABF2, have no carboxy-terminal domain at all (4,5,14). The proteins encoded by *T. thermophila* may have specialized functions associated with the unusual organisation of the genetic material in ciliate protozoans into micro- and

[1]The DNA sequence has been submitted to the GenBank database with the accession number M81360.

0006-291X/92 $1.50

macronuclei (16). *Babesia bovis* is a protozoan haemoparasite that does not exhibit this organisation of genetic material. Here we report the sequence of a cDNA clone encoding a *B. bovis* HMG1 homologue and compare the organisation of the protein with other HMG1 homologues.

MATERIALS AND METHODS

Bacterial strains. *Escherichia coli* strain Y1090 {Δ(*lac*U169), *pro*A+, Δ(lon), *ara*D139, *str*A, *sup*F, [*trp*C22::Tn10]}, (pMC9) was used as the host for the λgt11 constructs. *E. coli* strain JM109 {*end*A1, *rec*A1, *gyr*A96, *thi*, *hsd*R17(rk-, mk+), *rel*A1, *sup*E44, l-, Δ(*lac-pro*AB), [F', *tra*D36, *pro*AB, *lac*Iq, *lac*ZΔM15]} was used as the host for all of the plasmids constructed.

Library screening and DNA sequence and analysis. The λgt11 cDNA expression library constructed from *B. bovis* (Samford attenuated line) mRNA was a gift from K. R. Gale. To screen the library, 200,000 plaques were plated out and transfered to nitrocellulose membranes using standard procedures (17). The membranes were incubated with a bovine polyclonal antiserum (Pab8A6) raised against *B. bovis* proteins eluted from an affinity column carrying an anti-*B. bovis* monoclonal antibody (Mab) designated 8A6 (B. V. Goodger, unpublished). Positive plaques were identified and purified using standard methods (17). One recombinant phage, designated λgt11-8A6B, was analysed in more detail. The ~400bp inserted *Eco*RI fragment was subcloned into pGEM3Zf(+) (Promega Corporation) for sequence analysis. Both strands of the inserted fragment were sequenced by the dideoxy chain-termination method. The DNA and protein sequences were analysed using the MacVector 3.5 program (International Biotechnologies, Inc). DNA was labelled with [α^{32}P]dCTP using a Multiprime kit (Amersham Int. plc). Hybridization was carried out at 37°C in 5 x SSPE (20 x SSPE is 3.6M NaCl, 0.2M NaH_2PO_4, 0.02M EDTA, pH 7.4), 30% formamide, 5 x Denhardt's solution (0.1% ficoll, 0.1% polyvinylpyrrolidone, 0.1% bovine serum albumin), 0.5% SDS and 20μg denatured herring sperm DNA. Filters were washed in 2 x SSC (20 x SSC is 3M NaCl, 0.3M Na citrate) and 0.1% SDS at room temperature.

RESULTS AND DISCUSSION

The recombinant clone, λgt11-8A6B, expressed a β-Galactosidase (β-Gal) fusion protein of molecular weight ~130,000 (data not shown). The bovine polyclonal antiserum (Pab8A6) bound to the fusion protein, but not to β-Gal (data not shown). From the DNA sequence of the *Eco*RI insert in λgt11-8A6B an open reading frame in frame with β-Gal, and terminating prior to the 3' of the cDNA, was identified (Fig. 1). The product of this open reading frame had a calculated molecular weight of 13,581.

Identification of a putative HMG1 homologue. The predicted amino acid sequence of the open reading frame was used to search the translated GenBank

```
GCCGCCTATTGGGCATATTTTGGACGGGCTTGTTTTCTTTTTTGGTTCATTCTGGTCTCT 60
   R  L  L  G  I  F  W  T  G  L  F  S  F  L  V  H  S  G  L

TTGCTACGATGGCTGGTGCTTCGGACCGTACTGGCGTGCGCAGACCTAGGAAGGCTAAGA 120
F  A  T  M  A  G  A  S  D  R  T  G  V  R  R  P  R  K  A  K

AGGATCCCAACGCTCCTAAGCGTGCTTTATCTTCTTACATGTTTTTTGCTAAGGAGAAGC 180
K  D  P  N  A  P  K  R  A  L  S  S  Y  M  F  F  A  K  E  K

GTGTAGAGATAATTGCTGAGAACCCTGAGATCGCTAAGGACGTTGCTGCTATCGGTAAGA 240
R  V  E  I  I  A  E  N  P  E  I  A  K  D  V  A  A  I  G  K

TGATCGGTGCTGCTTGGAATGCTCTTTCTGATGAGGAGAAGAAGCCCTATGAGCGCATGT 300
M  I  G  A  A  W  N  A  L  S  D  E  E  K  K  P  Y  E  R  M

CTGATGAAGATCGTGTTAGGTACGAGCGTGAGAAGGCTGAGTACGCCCAGAGGAAGGTAT 360
S  D  E  D  R  V  R  Y  E  R  E  K  A  E  Y  A  Q  R  K  V

GATGCGAATTGTATTATTAATTTCTTTGGTTATCTAAAAAA                    401
```

Fig. 1. **DNA sequence of the *Eco*RI cDNA insert from λgt11-8A6B.** The predicted amino acid sequence is shown beneath the DNA sequence, with the proposed native protein indicated by bold type. Flanking *Eco*RI sites are not shown.

database (release 69). A number of protein sequences with significant similarities to the *B. bovis* amino acid sequence were identified. These proteins all contained HMG-Boxes. The *B. bovis* sequence contained fourteen of the sixteen highly conserved residues of HMG-Boxes (Fig. 2). The proteins with the highest similarity scores were the *S. cerevisiae* NHP6A and NHP6B proteins and HMG1 and HMG2 and their homologues (Fig. 2). Other proteins containing HMG-Boxes, including hUBF (6), SRY (7), mat1-M (11) and the mitochondrial proteins (4,5), had much lower similarity scores (data not shown).

On the basis of the sequence alignment to the HMG homologues the ATG at nucleotides 69-71 is proposed to be the normal translation start site in *B. bovis* (Fig. 1). It was the first ATG in the cDNA insert and an A at -3 is almost invariant in *B. bovis* genes and a G at -1 is very commom (B. P. Dalrymple, unpublished data). The amino acid sequence preceding this ATG did not exhibit any similarities to any HMG-Box containing proteins characterised so far. The protein encoded from this ATG was 97 amino acids long with a predicted molecular weight of 11,116.

The aligned *B. bovis* and *S. cerevisiae* NHP6A sequences had approximately 45% amino acid identity between the complete proteins, with approximately 57% identity in the HMG-Box (Fig. 2). Between amino acids 17 and 34 there was only one difference in the amino acid sequences of the two proteins (Fig. 2). This high level of homology was reflected in the DNA sequence with ~82% identity in the equivalent region of the gene (data not shown). This highly conserved amino acid sequence includes a proposed DNA-binding site (10,18), suggesting that the *B. bovis* and the *S. cerevisiae* proteins may exhibit very similar, if not identical, DNA binding properties. Also like the *S. cerevisae* proteins, the *B. bovis* protein did not have either an acidic or a basic carboxy-terminal tail.

The next most similar proteins were the metazoan HMG proteins. The HMG-Box in HNP1 exhibited ~45% amino acid identity with the HMG-Box in the *Zea.*

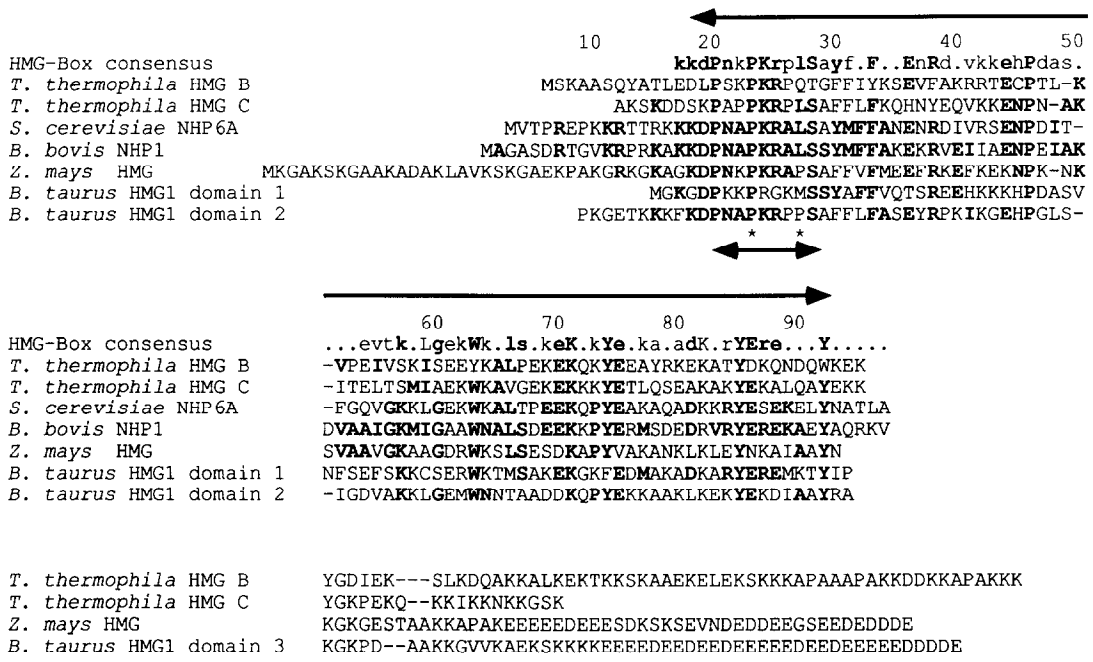

Fig. 2. **Alignment of the protein sequences of Non-Histone Proteins and HMG1 proteins**. Protein sequences were taken from the following sources; *T. thermophila* (13,21), *S. cerevisiae* (14), *Z. mays* (10) and *B. taurus* (22). Numbering is for the *B. bovis* sequence. Amino acids highlighted in bold type are identical between *B. bovis* and any of the other sequences. The HMG-Box consensus sequence is that proposed by Kolodrubetz (1), upper case letters indicate highly conserved amino acids, lower case letters indicate less well-conserved positions and "." indicates a nonconserved position. The arrows above the protein sequences indicate the extent of the HMG-Box (1), the small arrow below the sequence indicates a proposed DNA-binding domain (10); within this domain "*" indicates amino acid residues critical for the DNA binding of the HMG-Box containing protein SRY (18).

mays protein and 35-39% amino acid identity with the two HMG-Boxes in *Bos. taurus* HMG1. In contrast NHP1 was not very similar to the *T. thermophila* HMG homologues. It exhibited ~30% amino acid identity with the HMG-Boxes in the HMGB and HMGC. The greater similarity between the proteins from *B. bovis*, *S. cerevisiae* and the metazoans, than between the proteins from the two species of protozoa, suggests that the *T. thermophila* proteins may not be true HMG1 homologues. Alternatively, significant change in function (development of the micronuclear-macronuclear organisation for example) has led to a faster evolution of these proteins in the ciliate protozoa (see 15).

However, the protein encoded by the *B. bovis* cDNA clone appears to be a true homologue of *S. cerevisiae* NHP6A and NHP6B and we propose that the *B. bovis* protein be designated as non-histone protein 1 (HNP1). The cellular functions of these proteins are unknown. NHP6A and NHP6B are moderately abundant nuclear proteins that bind preferentially to single stranded DNA, but without an apparent sequence specificity (1,19). The difference in function of proteins with

different numbers of HMG-Boxes is not known, but it appears that a higher number of HMG-Boxes may confer a higher level of sequence specificity in the binding of the protein to DNA (see 4-11). The role of the positive and negatively charged carboxy-terminal domains is also unknown.

The HNP1 gene is not a member of a multigene family. The ~400bp *Eco*RI fragment was hybridized to *Eco*RI, *Hin*dIII, *Bgl*II and *Bam*HI digested*B. bovis* DNA, with all enzymes used except *Bam*HI only one hybridizing fragment was observed (data not shown). There is a *Bam*HI site in the sequenced gene. Thus, in contrast to most other characterised organisms, for which two or more closely related genes have been identified (2, 12, 16, 20), there is no evidence for a family of closely related NHP genes in *B. bovis*.

Acknowledgment: We would like to thank Dr. K. R. Gale for the gift of the *B. bovis* cDNA expression library.

REFERENCES

1. Kolodrubetz, D. (1990) Nucleic Acids Res. 18, 5565.
2. Wen, L., Huang, J-K., Johnson, B. H. and Reeck, G. R. (1989) Nucleic Acids Res. 17, 1197-1214.
3. Shirakawa, H., Tsuda, K. and Yoshida, M. (1990). Biochem. 29, 4419-4423.
4. Parisi, M. A. and Clayton, D. A. (1991) Science 252, 965-969.
5. Diffley, J. F. X. and Stillman, B. (1991). Proc. Natl. Acad. Sci. USA. 88, 7864-7868.
6. Jantzen, H-M., Admon, A., Bell, S. P. and Tjian, R. (1990) Nature 344, 830-836.
7. Sinclair, A. H., Berta, P., Palmer, M. S., Hawkins, J. R., Griffiths, B. L., Smith, M. J., Foster, J. W., Frischauf, A-M., Lovell-Badge, R. and Goodfellow, P. N. (1990) Nature 346, 240-244.
8. Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Munsterberg, A., Vivian, N., Goodfellow, P. and Lovell-Badge, R. (1990) Nature 346, 245-250.
9. Wetering, M. van de, Oosterwegel, M., Dooijes, D. and Clevers, H. (1991) EMBO J. 10, 123-132.
10. Bachvarov, D. and Moss, T. (1991) Nucleic Acids Res. 19, 2331-2335.
11. Kelly, M., Burke, J., Smith, M., Klar, A. and Beach, D. (1988) EMBO J. 7, 1537-1547.
12. Grasser, K. D. and Felix, G. (1991) Nucleic Acids Res. 19, 2573-2577.
13. Laux, T. and Goldberg, R. B. (1991) Nucleic Acids Res. 19, 4769.
14. Kolodrubetz, D. and Burgum, A. (1990) J. Biol. Chem. 265, 3234-3239.
15. Schulman, I. G., Wang, T., Wu, M., Bowen, J., Cook, R. G., Gorovsky, M. A. and Allis, C. D. (1991) Mol. Cell. Biol. 11, 166-174.
16. Schulman, I. G., Wang, T., Stargell, L. A., Gorovsky, M. A. and Allis, C. D. (1991) Dev. Biol. 143, 248-257.
17. Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular cloning. A laboratory manual. 2nd ed. Cold Spring Harbor Laboratory Press.
18. Nasrin, N., Buggs, C., Kong, X. F., Carnazza, J., Goebl, M. and Alexander-Bridges, M. (1991) Nature 354, 317-320.
19. Kolodrubetz, D. Haggren, W. and Burgum, A (1988) FEBS letts. 238, 175-179.
20. Pentecost, B. T., Wright, J. M. and Dixon, G. H. (1985) Nucleic Acids Res. 13, 4871-4888.
21. Roth, S. Y., Schulman, I. G., Cook, R. G. and Allis, C. D. (1987) Nucleic Acids Res. 19, 8112.
22. Kaplan, D. J. and Duncan, C. H. (1988) Nucleic Acids Res. 16, 10375.