Nucleic Acids Research, 1993, Vol. 21, No. 18 4369–4371

dHMG-Z, a second HMG-1-related protein in *Drosophila melanogaster*

Sarbjit S.Ner, Mair E.A.Churchill, M.Alexandra Searles and Andrew A.Travers
MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Received March 22, 1993; Revised and Accepted July 29, 1993 EMBL accession no. X71139

ABSTRACT

We report the identification of dHMG-Z, a gene related to dHMG-D and encoding a second invertebrate homologue of HMG 1 protein. The encoded proteins is 65% identical to dHMG-D protein, and also contains a single HMG-box as the DNA recognition motif. Analgous to dHMG-D, two transcripts are observed for dHMG-Z which are differentially regulated, and are the product of zygotic transcription unlike the dHMG-D transcripts which arise from both maternal and zygotic transcription. The genes for dHMG-D and dHMG-Z are located on adjacent loci in the genome and each contains two introns. The position of the second intron in the coding region is conserved between the two genes suggesting a common origin via gene duplication.

INTRODUCTION

Recently, a diverse group of proteins have been identified which function principally as transcription factors, and share a homologous region of 80 amino acids termed the 'HMG-box', with the HMG1/2 proteins (1). Some of the factors are necessary for genome maintenance (ABF2; 2) or are required for general transcriptional regulation (SIN1; 3), whereas others are involved in key cell differentiation pathways. These include, SRY, the testis determining factor (4), PolI transcription factor hUBF (5), V-(D)-J recombination recognition protein (6) and fungal mating-type proteins, Mc of *Schizosaccharomyces pombe* (7) and mat-1 of *Neurospora crassa* (8). The HMG-box is a DNA-binding motif for which a NMR solution structure was recently reported (9). It can recognize specific DNA sequences (10–13), or bind DNA molecules containing particular structures, such as cruciforms (14) and bent or kinked DNA (15; 16). In addition some HMG-box containing proteins have the capacity to bend or loop DNA (17). The HMG-box is thus a versatile DNA-binding motif which is used in different structural and functional contexts.

We report here the cloning and genomic organization of dHMG-Z, a second *Drosophila* homologue of the HMG-1/2 type. This protein is closely related to dHMG-D—which had been reported (1, 18) but was recently isolated and purified by Elgin and co-workers (19). The two proteins (dHMG-D and dHMG-Z) are highly homologous to similar proteins in another Dipteran *Chironomus thummi* (20). Northern hybridization and primer extension analyses reveal that transcripts for these proteins are relatively abundant and expressed during all stages of early embryogenesis. However whereas dHMG-D is transcribed both maternally and in the zygote, dHMG-Z RNA is principally zygotic. There are families of transcripts associated for each gene reflecting a dynamic tissue-specific expression and a complex mode of regulation of the two loci (SSN, unpublished).

RESULTS

Drosophila DNA encodes two highly related HMG 1/2-like proteins

dHMG-D was initially isolated by screening a 0–16 hr *Drosophila* embryonic cDNA library cloned in λgt11 expression vector using complementary oligonucleotides, corresponding to nucleotides 415–454 of the *ftz* USE which are protected by a 0–4 hr but not by a 5–10 hr embryonic extract (22). The comparison of sequences from the lambda clone and the full length cDNA clones indicated a discrepancy in the 5′ end of the clones (data not shown) that led us to adopt a PCR based strategy to identify additional related sequences (described in Materials and Methods). This analysis resulted in the isolation of two cDNA clones (1.128 kb and 0.64 kb) encoding a homologous protein termed dHMG-Z to dHMG-D. The predicted protein is 111 amino acids and is 65% identical and 78% similar at the conserved amino acid level to dHMG-D.

dHMG-Z and dHMG-D lie on adjacent loci

The 1.128 kb fragment containing dHMG-Z sequences was used to screen a *Drosophila* genomic library and two overlapping phage clones isolated (λEmbl3-Z1, λEmbl3-Z2). Phage clones were also isolated using dHMG-D specific sequences (λEmbl3-N1, λEmbl3-N2). Phage λEmbl3-N2 overlaps with λEmbl3-Z1 suggesting that dHMG-D and dHMG-Z are located on overlapping inserts and hence on adjacent genetic loci. We have characterized the insert in λEmbl3-N2 by restriction and sequence analyses and PCR using specific oligodeoxynucleotide primers and observe it to contain all exons and introns for dHMG-D but only the first exon and part of the first intron for dHMG-Z. This data is consistent with an head-to-head arrangement for the two genes (Fig. 1). This analysis reveals the dHMG-D gene contains two introns. The first is within the 5′ UTR which results in a short exon 1 (30 bp) followed by a 1.2 kb intervening sequence (IVS1). The second exon is 215 bp and encodes for the N-terminal 39 residues. IVS2 is short (65 bp) and exon3 contains the rest of the coding region and the untranslated 3′ UTR (1.030 kb hmgd1 and 780 bp hmgd2 in two cDNA clones isolated

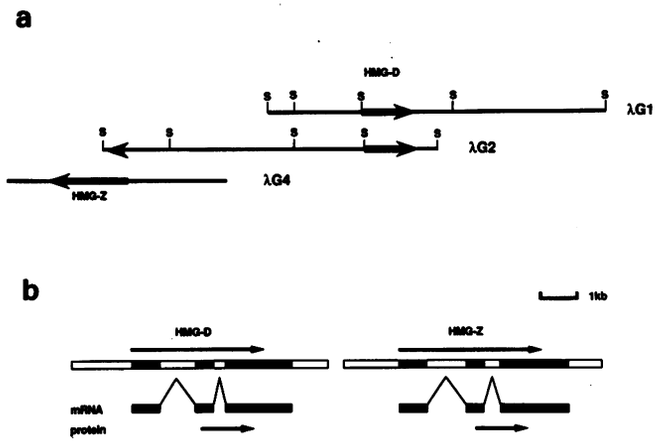

Figure 1. Genomic organization of dHMG-D and dHMG-Z. **a.** Three overlapping phage clones isolated by screening Embl3 library with sequences unique to dHMG-D and dHMG-Z and subsequent mapping reveal the genes are within the same chromosomal location and organised in a head-to-head arrangement. λG2 contains the entire coding sequences of dHMG-D and the first exon of dHMG-Z. S, SalI. **b.** Intron/exon orgainzation of dHMG-D and dHMG-Z as inferred from sequence analysis of genomic and cDNA fragments of both genes. Exons are represented by filled boxes.

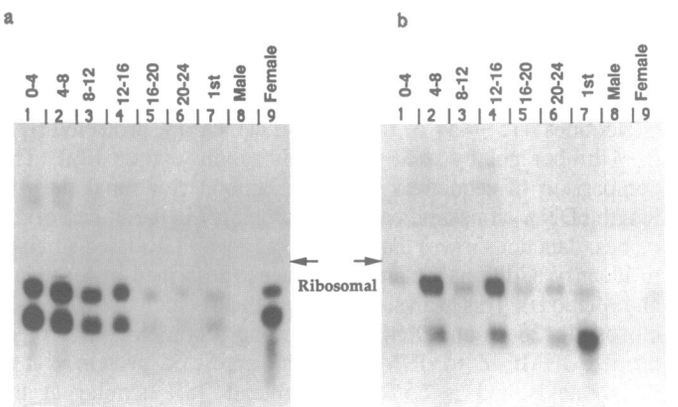

Figure 2. a. Northern blot analysis of total mRNA prepared from staged embryos, 1st instar larvae, 2nd instar larvae and female flies hybridized with 1.073 kb *Eco*RI fragment from lambda phage λ1α12. 20 μg of total RNA were separated on a 1.2% agarose gel containing 0.35M formaldehyde and transferred to nitrocellulose membrane. The probe detects two major transcripts of ~1 kb and ~1.5 kb and two minor transcripts at approximately 3 kb. Lanes: 1, RNA from 0–4 hr embryos, 2. 4–8 hr embryos, 3. 8–12 hr embryos, 4. 12–16 hr embryos, 5. 16–20 hr embryos, 6. 20–24 hr embryos, 7. 1st instar larvae, 8. adult male, 9. adult female flies. **b.** Northern analysis of dHMG-Z specific transcripts. A riboprobe of 5′ sequences unique to dHMG-Z (from hmgz1, Fig. 1b) was used to probe total RNA from Lanes: 1. 0–4 hr embryos, 2. 4–8 hr embryos, 3. 8–12 hr embryos, 4. 12–16 hr embryos, 5. 16–20 hr embryos, 6. 20–24 hr embryos, 7. 1st instar larvae, 8. adult male, 9. adult female flies. Two dHMG-Z species are identified in RNA from embryos. Total RNA from 8–12 hr embryos (lane 3) is under represented and hence the dHMG-D and dHMG-Z transcripts appear less abundant than in lane 2 and lane 4.

for dHMG-D). Similarly the intron exon boundaries were determined for the dHMG-Z locus revealing three exons (exon1 165 bp, exon2 214 bp and exon3 741 bp) separated by two introns (IVS1 is over 1 kb (the exact length has not been determined) and IVS2 is 95 bp). This indicates that the exon/intron positions are conserved relative to dHMG-D. In particular the precise insertion of the second intron is identical for both genes suggesting a common origin for the two genes.

dHMG-Z is primarily a zygotic transcript present throughout embryogenesis

We investigated the expression pattern of the dHMG-Z transcripts (Fig. 2) using sequences from hmgz1, the larger of the two cDNA clones. Two major transcripts (approx. 0.65kb and 1.2kb) are observed. The lane containing RNA from 0–4 hr embryos shows a small amount of the 1.2 kb transcript suggesting that either the maternal contribution of dHMG-Z is minor or this species arises from early zygotic transcription. 4–8 hr and 8–12 hr embryos show both the major transcripts with the 1.2 kb message being more abundant. Both the transcripts are present in later embryos but at reduced levels. RNA from 1st instar larvae reveals relatively high level of the smaller message than from 12–16 or 16–20 hr embryos suggesting a requirement for the dHMG-Z protein during larval development. Interestingly, a pair of larger transcripts similar to those corresponding to the 3–4 kb messages observed with dHMG-D specific probes is also apparent. Compared to dHMG-D transcripts dHMG-Z transcripts are overall less abundant (as observed from Northern and *in situ* analysis).

Northern analysis of dHMG-D also reveals two transcripts (approx. 0.9kb and 1.3 kb) present at relatively high levels in RNA prepared from 0–4 and 4–8 hr embryos with slightly lower amounts present in the 8–12 and 12–16 hr embryos. Trace amounts of transcripts are present in later stages of development. In the female adult fly, but not in the male, there is again a relatively high level of both transcripts. However, the level of the larger transcript is only one fifth that of the smaller transcript suggesting the transcripts are differentially regulated and that the smaller transcript constitutes the principal maternal message. We believe that the two isolated cDNA clones correspond to these two transcripts.

DISCUSSION

We have isolated cDNA clones encoding a second *Drosophila* protein related to HMG 1/2 called dHMG-Z that is only expressed in the zygote. Based on their relative abundance, association with chromatin (19), and the high level of homology at the amino acid level, we suggest dHMG-D and dHMG-Z are invertebrate homologues of HMG 1/2 proteins but containing only a single HMG-box.

The HMG-box of dHMG-Z, which falls into the group of the HMG-1/2 type, is likely to be involved in recognition of the DNA (12; 13), and differs from the HMG-box of the transcription factors in the pattern of basic residues and prolines in the N-terminus (1). These residues appear to be important for DNA recognition in SRY and IRE-ABP where mutagenesis of residues 4 and 8 altered the specificity of the DNA binding interaction (12), and therefore the sequence of residues 3–6 of dHMG-Z (RPKR) which differs from that of dHMG-D (KPKR) in this region may also modulate the affinity or selectivity of dHMG-D and dHMG-Z for DNA and hence dHMG-Z may have a subtly different biological role from dHMG-D protein.

The reason for the presence of two HMG 1/2 related proteins in Drosophila is unknown. Two genes encoding homologous proteins to dHMG-Z and dHMG-D, have been identified in the insect *Chironomus thummi* (20) and hence this may be a general

feature in dipterans. It is also interesting that the yeast *Saccharomyces cerevisiae* also contains two highly related HMG-1/2 proteins, NHP6a and NHP6b (23) as does *Tetrahymena* (HMG-B and HMG-C). The reasons for this may be similar to why vertebrates have two very homologous proteins HMG 1 and HMG 2. The developmental expression profiles of the *Drosophila* proteins may suggest that the properties of dHMG-D, the mRNA for which is deposited in the egg by the mother, are more compatible with rapid cell division cycles which occur prior to cellularization and that dHMG-Z fulfills a similar function during later stages of development.

MATERIALS AND METHODS

Screening expression library

The denaturation−renaturation protocol described by Vinson *et al.* (21) was used to screen a 0−16hr embryonic cDNA library cloned into λgt11 expression vector. The probe was an end labelled dimer sequence of the # 12 binding site corresponding to nucleotides 421−441 of the USE element (22).

Isolation and sequence analysis of dHMG-D and dHMG-Z

The predicted polypeptides for the lambda and the cDNA clones differ at the N-termini. The lambda clone predicts an additional 39 amino acids. Interestingly, 66 nucleotides of this stretch is present in cDNA hmgd2 in the reverse orientation. We are unable to explain this anomaly and suggest that it may reflect a cloning artifact. However, 22 nucleotides remain unique to the lambda clone. We wondered if the entire stretch of nucleotides (22 unique and 66 in the reverse orientation) arose from a further cDNA clone which contained additional 5′ sequences, since the Northern data shows the presence of a pair of transcripts which are larger and less abundant than the major transcripts. The initial screening of the embryonic plasmid library with the lambda insert resulted in over 300 positives. Further analysis of a subset of these positives revealed only sequences corresponding to the cDNAs already isolated, possibly because the larger transcripts are less abundant. Therefore a strategy based on the polymerase chain reaction was employed to attempt to clone these larger transcripts. To amplify and sub clone sequences 5′ to the coding region a complementary oligonucleotide corresponding to nucleotides encoding residues 12−19 (SN147, 5′- GGGCACTGTTGAG-CCACAGCATGTA-3′) was used in a PCR reaction together with an oligonucleotide which hybridized to a sequence in the plasmid used in the construction of the cDNA library. In addition to sequences corresponding to cDNAs hmgd1 and hmgd2 we observed two clones (x-16 and x-31) which differed in their 5′ nucleotide sequence. Since the oligonucleotide used in the PCR reaction hybridized to sequences encoding amino acids 12−19, both x-16 and x-31 contain sequences coding for the N-terminal 11 amino acids. These, however, were not identical to those observed with cDNA hmgd1 and hmgd2. There is an insertion of a glycine codon at residue 2 and a lysine to arginine change at residue 4. Therefore, we suggest that a second protein related to dHMG-D is present in Drosophila. To isolate full length cDNA clones encoding the dHMG-Z protein, a probe derived by PCR using oligonucleotides SNX5 (5′-CCCAGTACATTTTGTTCA-G-3′) and SNX6 (5′-TTTGCCTTGTCGCTGTGCTTCTG-3′) with clone x-31 was used to screen a 4−8 hr library. From over 100 positives approximately 20 were analyzed further and revealed two sizes of insert (0.65 kb and 1.2 kb). We have termed this gene dHMG-Z. cDNA clones containing full length dHMG-Z sequences were subsequently isolated using the 5′ UTR sequence as a probe to screen a 8−12hr cDNA library. This resulted in two sizes of insert (0.65 kb (hmgz2) and 1.2 kb (hmgz1)). The larger contained a longer 3′ UTR.

ACKNOWLEDGEMENT

This work was supported by the Medical Research Council, UK.

REFERENCES

1. Ner, S.S. 1992. *Current Biology* **2**:208−210.
2. Diffley, J.F.X. and B. Stillman. 1991. *Proc. Natl. Acad. Sci. USA* **88**:7864−7868.
3. Kruger, W. and I. Herskowitz. 1991. *Mol. Cell. Biol.* **11**:4135−4146.
4. Sinclair, A.H., P. Berta, M.S. Palmer, J.R. Hawkins, B.L. Griffiths, M.J. Smith, J.W. Foster, A-M. Frischauf, R. Lovell-badge and P.N Goodfellow. 1990. *Nature* **346**: 240−244.
5. Jantzen, H-M., A. Admon, S.P. Bell and R. Tjian. 1990. *Nature* **344:** 830−836.
6. Shirakata, M., K. Huppi, S. Usuda, K. Yoshida and H. Sakano. 1991. *Mol. Cell. Biol.* **11**: 4528−4536.
7. Kelly, M., J. Burke, M. Smith, A. Klar and D. Beach. 1988. *EMBO J.* **7**: 1537−1547.
8. Staben, C. and C. Yanofsky. 1990. *Proc. Natl. Acad. Sci.* **87**: 4917−4921.
9. Weir, H.M., P.J. Kraulis, C.S. Hill, A.R.C. Raine, E.D. Laue and J.O. Thomas. (1993). *EMBO J.* **12**:1311−1319.
10. van de Wetering, M., M. Oosterwegel, D. Dooijes and H. Clevers. 1991. *EMBO J.* **10**: 123−132.
11. Bianchi, M.E., L. Falciola, S. Ferrari and D.M.J. Lilley. 1992. *EMBO J.* **11**:1055−1063.
12. Nasrin, N. C. Buggs, X.F. Kong, J. Carnazza, M. Goebl and M. Alexander-Bridges. 1991. *Nature* **354**:317−320.
13. Harley, V.R., D.I. Jackson, P.J. Hextall, J.R. Hawkins, G.D. Berkowitz, S. Sockanathan, R. Lovell-badge and P.N. Goodfellow. 1992. *Science* **255**:453−456
14. Bianchi, M.E., M. Beltrame and G. Paonessa. 1989. *Science* **243:** 1056−1059.
15. Bruhn, S.L., P.M. Pil, J.M. Essigmann, D. E. Housman and S. J. Lippard. 1992. *Proc. Natl. Acad. Sci.* **89**: 2307−2311.
16. Ferrari, S., V.R. Harley, A. Pontiggia, P.N. Goodfellow, R. Lovell-Badge and M. E. Bianchi. 1992. *EMBO J.* **11**:4496−4506.
17. Giese, K., Cox, J. and Grosschedl, R. 1992. *Cell* **69**, 185−195.
18. Bassuk, J.A. and J. E., Mayfield. 1982. *Biochemistry* **21**: 1024−1027.
19. Wagner, C.R., K. Hamana and S.C.R. Elgin. 1992. *Mol. Cell. Biol.* **12**:1915−1923
20. Wisniewski J.R and E. Schulze. 1992. *J. Biol. Chem.* **267**:17170−17177.
21. Vinson, C.R., K.L. LaMarco, P.F. Johnson, W.H. Landshultz and S.L. McKnight. 1988. *Genes Dev.* **2:** 801−807.
22. Harrison, S.D. and A.A. Travers. 1988. *Nucleic Acids Res.* **16:** 11403−11416.
23. Kolodrubetz, D. and A. Burgum. 1990. *J. Biol. Chem.* **265**:3234−3239.