Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 2414–2418, March 1993
Biochemistry

A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI2-like helicase domain

(transcription factor/chromatin/alternative mRNA processing/cDNA)

VÉRONIQUE DELMAS*, DAVID G. STOKES, AND ROBERT P. PERRY†

Institute for Cancer Research, Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, PA 19111

Contributed by Robert P. Perry, December 22, 1992

ABSTRACT **Two overlapping cDNAs that encode a 197-kDa sequence-selective DNA-binding protein were isolated from libraries derived from mouse lymphoid cell mRNA. In addition to a DNA-binding domain, the protein contains both a chromodomain, which occurs in proteins that are implicated in chromatin compaction, and an SNF2/SWI2-like helicase domain, which occurs in proteins that are believed to activate transcription by counteracting the repressive effects of chromatin structure. A Southern blot analysis indicated that this protein, which we have named CHD-1, for *c*hromodomain–*h*elicase–*D*NA-binding protein, is present in most, if not all, mammalian species. A Northern blot analysis revealed multiple CHD mRNA components that differed both qualitatively and quantitatively among various cell types. The various mRNAs, which are probably produced by alternative RNA processing, could conceivably encode tissue-specific and developmental stage-specific isoforms of the protein. Based on its interesting combination of features, we suspect that CHD-1 plays an important role in gene regulation.**

In eukaryotic cells, selective transcription of discrete sets of genes is accomplished by a variety of regulatory mechanisms. One major mechanism involves the concerted action of distinct combinations of proteins (transcription factors) that bind directly or indirectly to particular DNA sequences and act to stimulate or repress the basal transcriptional machinery (see ref. 1 for review). Another important mechanism regulates the accessibility of certain genes to the transcriptional apparatus by regional modifications of chromatin structure. Proteins participating in this type of mechanism could affect such processes as nucleosome packing, loop formation, DNA supercoiling, and attachment to the nuclear matrix (see ref. 2 for review). Our knowledge of transcription factors far exceeds that of the proteins involved in accessibility mechanisms.

Recently, genetic studies with yeast and *Drosophila* have identified an interesting class of proteins that appear to be involved in transcriptional regulation via their effects on chromatin organization (reviewed in ref. 3). In yeast, these proteins, known as SNF2/SWI2, SNF5, SNF6, SWI1, and SWI3, are required for the transcriptional activation of a large set of diversely regulated genes. Mutations in genes that encode histones and other chromatin-associated proteins can partially alleviate this requirement. In *Drosophila*, proteins such as Brahma (Brm) may play an analogous role. Brm apparently helps to counteract the repressive effect of the Polycomb (Pc) protein, which is believed to promote locus-specific chromatin compaction (4, 5). SNF2 and Brm are strikingly similar. They are of comparable size (194 vs. 185 kDa) and are 57% identical over a 630-residue region that contains a pattern of motifs typical of certain helicases (6). Neither of these proteins has yet been found to possess DNA-binding capability.

We report here on the isolation and characterization of cDNAs that encode a 197-kDa mouse protein with an SNF2 (Brm) pattern of helicase motifs in its central region and two other noteworthy features: a sequence-selective DNA-binding domain and a motif known as the chromodomain, which occurs in Pc and in HP1, a structural component of compacted chromatin (5, 21). The chromodomain is essential for the function of Pc, apparently being required for its assembly into chromatin as part of a multiprotein complex (5). Thus, the mouse protein, which we have termed CHD-1,‡ embodies within a single molecule characteristics of two functionally related proteins (Brm and Pc), as well as DNA-binding capability. This combination of features could endow CHD-1 with novel regulatory properties.

MATERIALS AND METHODS

Isolation of the κY3 and κY9 Clones and Plasmid Construction. λgt11 phage from a mouse B-cell lymphoma (A20) cDNA expression library (Clontech) were screened with double-stranded oligonucleotide probes by the protocol of Vinson *et al.* (7) with the modification of Beckman *et al.* (8). The initial probe consisted of a multimer of an oligonucleotide representing the −97 to −76 region of the V_κ19A promoter (κY-WT, Table 1) joined by 5′ *Xba* I and 3′ *Bam*HI linkers. After the third plaque purification, two λ phages, κY3 and κY9, were selected for further analysis. *Eco*RI fragments containing the cDNA sequences were excised from the selected phages and inserted into the *Eco*RI site of either the pGEM-4Z vector (Promega) or the maltose fusion protein bacterial expression vector pMAL-CRI (Biolabs; Northbrook, IL). The *Eco*RV and *Hin*dIII truncation mutants were constructed by standard procedures.

Isolation of Clone 3-3. A cDNA library was constructed from cytoplasmic poly(A)$^+$ RNA of S194 plasmacytoma cells and the λZap II vector (Stratagene). Briefly, cDNA was synthesized with a kit from Pharmacia and a 20-mer antisense primer (ATATAACAACGGTGTCAGCA) located ≈300 bp from the 5′ end of κY3 (Fig. 1). The library was subjected to one round of amplification and then screened with a 654-bp *Eco*RI–*Eco*RV fragment encompassing the 5′ end of κY3. Positive plaques were subjected to two more rounds of plaque purification and pBluescript plasmids containing the cDNA inserts were excised from the λZap II vector according to the manufacturer (Stratagene). One of the largest clones (3-3) was selected for sequencing on the basis of restriction digest and Northern blot analyses.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*Present address: Laboratoire de Génétique des Eucaryotes, Institut de Chemie Biologique, 11 rue Humann, 67085 Strasbourg, France.
†To whom reprint requests should be addressed.
‡The sequence reported in this paper has been deposited in the GenBank data base (accession no. L10410).

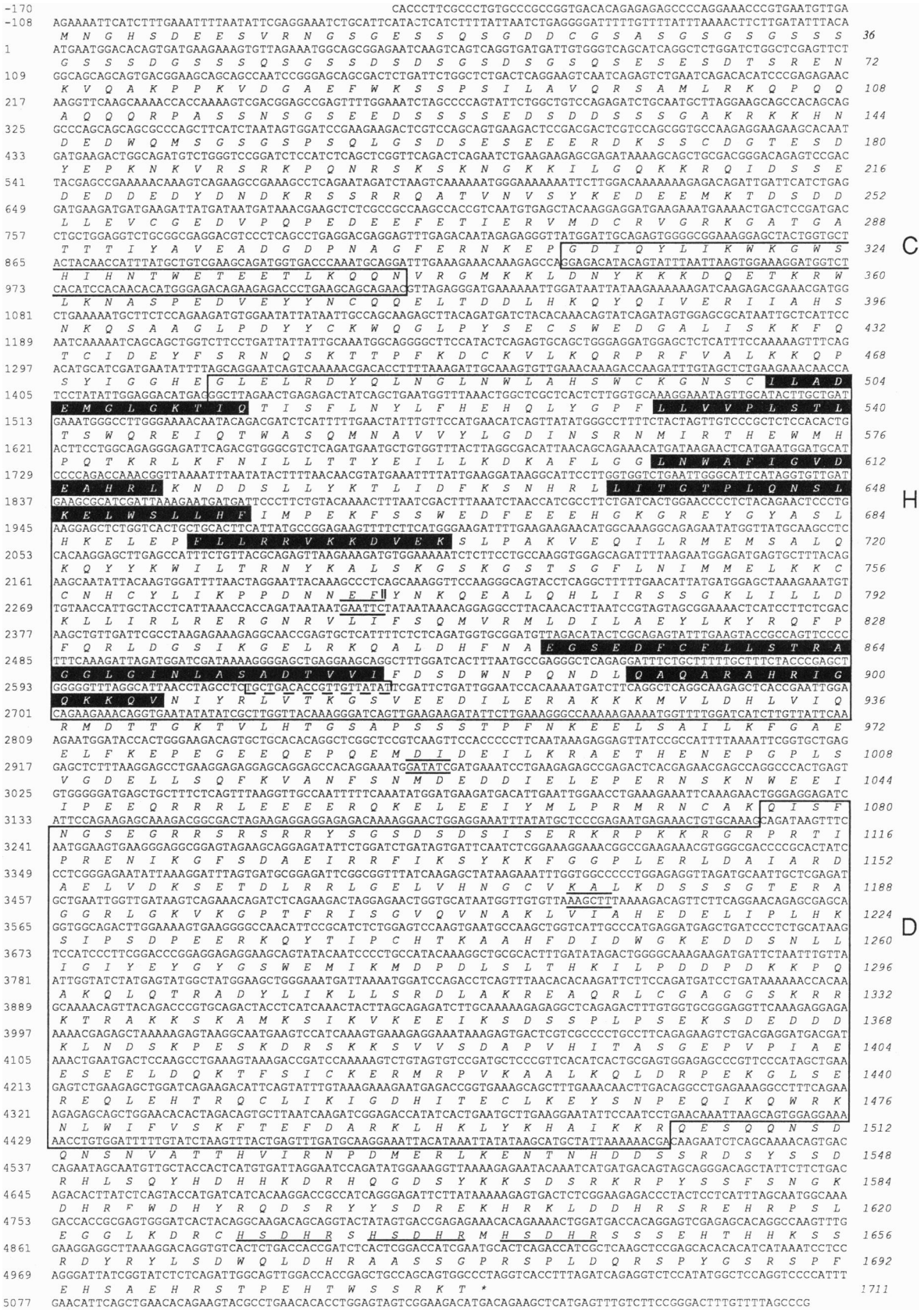

FIG. 1. Nucleotide sequence of CHD-1 cDNA and its deduced amino acid sequence. Numbers at left and right refer to adjacent nucleotides and amino acids, respectively. The boxed sequences indicate the chromodomain (C), SNF2/SWI2-like helicase domain (H), and the region containing the DNA-binding domain (D). The highly conserved helicase motifs are highlighted. The 5′ boundary of the κY3 and κY9 clones is indicated by a double vertical line (before nucleotide 2314). The 3′ boundary of the κY9 clone is at position 4512. The *Eco*RI (2308), *Eco*RV (2965), and *Hin*dIII (3527) sites are over- and underlined. The sequence of the primer used to generate clone 3-3 is indicated by a dashed underline. A basic motif repeated three times is underlined.

Sequencing. Chain-termination sequencing was carried out with the United States Biochemical Sequenase protocol employing dITP and 7-deaza-GTP to eliminate electrophoretic band compressions. Both strands of the 2865-bp κY3 insert and the 2874-bp 3-3 insert were sequenced by a combination of synthetic primers, subclones, and an exonuclease III deletion kit from Promega.

Analysis of Maltose-Binding Fusion Proteins. Maltose-binding fusion proteins were induced by isopropyl β-D-thiogalactopyranoside (0.3 mM) for 2 hr. The pelleted bacteria were suspended directly in sodium dodecyl sulfate/polyacrylamide gel electrophoresis (SDS/PAGE) sample buffer (9) and subjected to SDS/PAGE. Proteins were electroblotted onto nitrocellulose filters and incubated with monomeric wild-type and mutant κY probes under conditions identical to those used for the bacteriophage plaque screen.

Southern and Northern Blots. DNA and cytoplasmic poly(A)$^+$ RNA were extracted as described (10). Southern blots of *Eco*RI-digested DNA were made with Hybond N+ sheets (Amersham). The probe was hybridized for about 16 hr at 65°C in 10% SDS (which produces a final sodium concentration of 0.35 M) containing 7% polyethylene glycol and 7.5 mg of salmon sperm DNA per ml. Blots were washed twice for 15 min at room temperature in 2× SSC (0.3 M sodium chloride/0.03 M sodium citrate, pH 7)/0.1% SDS, then for 30 min at 55°C in 2× SSC and finally for 10 min at room temperature in 2× SSC. Northern blots were prepared and analyzed according to the manufacturer's protocols (Nytran; Schleicher & Schuell).

RESULTS

The project that led to the characterization of CHD-1 began as an attempt to clone a cDNA for κY, a nuclear protein that binds to certain immunoglobulin promoters and helps compensate for weak Oct factor interactions (11). Using a multimerized 27-bp probe that encompasses the κY binding site, we screened a λgt11 cDNA expression library for DNA-binding capability. From a plating of about 5×10^5 phage, we obtained two plaques that continued to score positive upon further purification. These recombinant phage (κY3 and κY9) were then subjected to a second screen with a battery of monomeric probes representing the wild-type κY sequence and four mutants: m4, m11, m5R, and m5L (Table 1). Both recombinants gave positive signals with the wild-type probe and the m4 mutant. No signals were detected with the other three mutant probes, indicating that the recombinant protein(s) selectively recognize only certain DNA sequences. However, the sequence specificity indicated by the plaque assays was not identical to the specificity of κY, determined by electrophoretic mobility-shift assays with a lymphoid cell nuclear extract and the same set of probes (Table 1). This discordance raised some uncertainty about whether the cloned cDNAs actually encoded κY. Nevertheless, we were sufficiently intrigued by the discriminating DNA-binding properties of the recombinant protein to proceed with an analysis of the cDNA sequences.

Table 1. Relative binding of κY wild-type (WT) and mutant probes to proteins in recombinant phage plaques and mouse nuclear extract

Probe	Sequence*	A+T, %	Binding	
			Plaques	Extract
κY-WT	ATAATTTACTTCCTTATTTGATG	78	+	+
m4	----------CGAG----------	74	+	−
m11	------GGCAGAGTGCG--------	57	−	−
m5L	-GCCAG----------	61	−	−
m5R	----------GGGCC—	61	−	+

The nuclear extract assay consisted of an electrophoretic mobility-shift analysis as described (11).
*Dashes in mutant sequences indicate identity with the wild type.

A preliminary restriction analysis and partial sequencing indicated that κY3 (2.9 kb) and κY9 (2.2 kb) are virtually identical over a 2.2-kb region that includes their 5′ ends (determined by orientation with respect to the λgt11 expression vector). The entire sequence of κY3 revealed a long open reading frame starting at the 5′ end and continuing for 940 codons. The absence of an in-frame methionine codon near the 5′ end indicated that sequences encoding the N-terminal portion of the protein were not present in these cDNAs. The lack of a recognizable AATAAA motif at the 3′ end of κY3 indicates that a portion of the 3′ untranslated sequence is also missing.

When the amino acid sequence corresponding to the κY3 open reading frame was used to search the GenBank and EMBL data banks (January, 1991), we observed that the initial 165 amino acids were 54% identical to a portion of the yeast SNF2/SWI2 protein (12). This striking finding impelled us to determine the complete sequence of the mouse protein. For this purpose, we produced a cDNA library specifically enriched in mRNA sequences that extend 5′-ward from κY3/κY9. This library, constructed with cDNA primed by an antisense oligonucleotide representing positions 305–324 of κY3, was screened with a 654-bp probe derived from the 5′ end of κY3. A positive clone that appeared to have one of the largest inserts (3-3) was sequenced and found to overlap the 5′ end of κY3, up to the primer sequence.

The composite sequence of clones κY3 and 3-3 contains 5349 bp and an open reading frame encoding 1711 amino acids (Fig. 1). The presumptive translational start codon, which resides in a sequence that conforms to the Kozak consensus (13) at key positions, is preceded by termination codons in all three reading frames. The region closely related to SNF2/SWI2 is located between amino acids 476 and 936. This segment contains the entire set of seven motifs that are conserved among certain families of helicases, including the presumptive nucleotide binding site (6). The spacing between these motifs is basically the same as that in SNF2/SWI2 and other recently characterized members of this family such as STH1 (6), Brm (4), and hSNF2L (14). In this putative helicase domain, the mouse protein is 47–49% identical and 67–69% similar to these four proteins over 461 amino acids.

Two other interesting characteristics of the mouse protein were noted: a 30-residue segment known as the chromodomain (15) at amino acids 312–341 and a thrice-repeated pentamer, His-Ser-Asp-His-Arg, between amino acids 1629 and 1645. The chromodomain motif has been identified in proteins that are believed to be implicated in chromatin compaction (5, 21). A comparison of this chromodomain sequence with the chromodomain sequences of five other proteins (Fig. 2) shows that it contains almost all of the highly conserved residues. The significance of the pentamer repeat is presently unknown. Based on its combination of features,

```
      1                  10                  20                  30
      *  *  * * * * * * * *   * * * * *  * * * *   *  *
CHD-1 G D I Q Y L I K W K G W S H I H N T W E T E E T L K Q Q N V
M3    G K L E Y L V K W R G W S S K H N S W E P E E N I L D P R L
Pc    G V V E Y R V K W K G W N Q R Y N T W E P E V N I L D R R L
M1    G K V E Y L L K W K G F S D E D N T W E P E E N L D C P D L
M2    G K V E Y F L K W K G F T D A D N T W E P E E N L D C P E L
HP1   G K V E Y Y L K W K G Y P E T E N T W E P E N N L D C Q D L
```

FIG. 2. Comparison of the chromodomain segment of CHD-1 with the corresponding segments of five other proteins. HP1, heterochromatin-specific protein (21); Pc, Polycomb protein (15); M1 and M2, mouse modifier proteins 1 and 2 (16); and M3, mouse modifier protein 3 (17). Asterisks indicate positions at which CHD-1 is either identical or similar to at least one other member of the family.

we named this protein CHD-1 (for *c*hromodomain–*h*elicase–*D*NA-binding protein).

Since none of the other SNF2-related proteins have been observed to bind directly to DNA, it was important to verify that CHD-1 actually has sequence-selective DNA-binding capability and, if possible, to localize the domain responsible for this property. For this purpose we used the pMAL-cRI vector system (18) to direct the synthesis in *Escherichia coli* of a series of inducible fusion proteins consisting of the maltose-binding protein joined to various portions of CHD-1. Based on the results of the original plaque screen, we expected that the DNA-binding domain would be located within the region common to κY3 and κY9. Accordingly, we constructed recombinant plasmids containing the entire κY3 or κY9 coding sequence or truncated portions that extend from the 5′ end of κY3/κY9 (amino acid 772) to a *Hin*dIII site at amino acid 1176 or an *Eco*RV site at amino acid 989. Although the relatively large fusion proteins produced by the κY3 and κY9 plasmids (about 145 and 122 kDa, respectively) were readily detected in bacterial lysates when analyzed by SDS/PAGE, they could not be solubilized by procedures that would allow their subsequent purification by affinity chromatography on amylose columns. Therefore, we size fractionated total bacterial lysates by SDS/PAGE, transferred the proteins to nitrocellulose, and examined their ability, after renaturation, to bind the wild-type κY oligonucleotide probe. Binding was observed with the κY3- and κY9-encoded proteins, but not with the *Hin*dIII- or *Eco*RV-truncated derivatives (Fig. 3*A*). When the κY9 fusion protein was tested with the battery of mutant probes, large differences in binding capacity were observed (Fig. 3*B*). This test, which was designed to be more quantitative than the plaque assays, indicated that the wild-type sequence was preferred to that of m4 and that there was detectable, albeit weak, binding of the m5L and m5R probes. Discrimination between the wild-type and m11 probes was at least 25-fold. These results confirm that CHD-1 binds to DNA in a sequence-selective manner. Moreover, they show that a portion of the protein between residues 1176 (*Hin*dIII site) and 1504 (end of the κY9 cDNA) is required for DNA binding. A more extensive analysis with additional constructs (data not shown) indicated that the DNA-binding domain lies within the region bounded by amino acids 1077 and 1504 (Fig. 1).

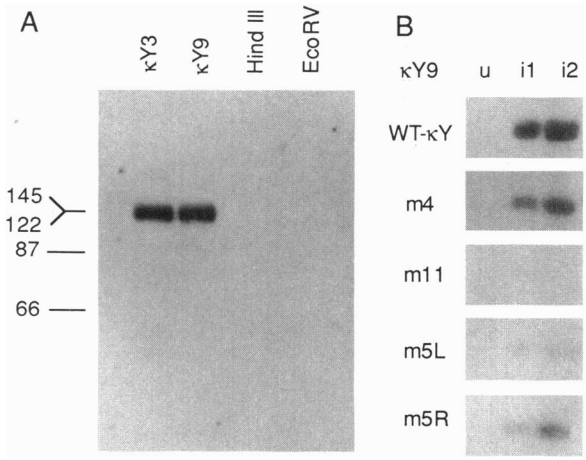

FIG. 3. Localization of the CHD-1 DNA-binding domain and test of sequence selectivity. (*A*) κY3 and κY9 cDNAs and two 3′ deletion mutants truncated at the *Hin*dIII or the *Eco*RV sites (Fig. 1) were cloned into the *Eco*RI site of pMAL-CRI, which results in the fusion of the cDNA-encoded protein to the maltose-binding protein. Plasmid expression was induced by isopropyl β-D-thiogalactopyranoside, and the proteins in total cell lysates were analyzed as described in *Materials and Methods*. The approximate sizes (in kilodaltons) of the fusion proteins detected by Coomassie staining and compared with a set of reference markers are indicated at left. (*B*) Lysates from uninduced (u) or induced (i) bacteria containing the pMAL-CRI-κY9 expression plasmid were analyzed as in *A* with the wild-type and mutant probes listed in Table 1. i1 and i2 represent single and double amounts of induced cell lysate, respectively.

To determine whether CHD-1 is highly conserved among mammals, we probed a Southern blot of *Eco*RI-digested genomic DNA from diverse mammalian sources with a 654-bp *Eco*RI–*Eco*RV fragment that encompasses a portion of the helicase domain. The results (Fig. 4) indicated that genes encoding CHD-1 homologues are present in most, if not all, mammalian species. The simplicity of the Southern blot patterns indicates that there is only one or a few such genes in each genome. Because of extensive mismatches, the genes encoding homologues of SNF2, STH1, Brm, and hSNF2L would probably not be detected by this analysis. For example, the 40% interspersed mismatches between the DNA sequences encoding this region of CHD-1 and hSNF2L should preclude the formation of stable DNA duplexes.

Northern blots of cytoplasmic poly(A)$^+$ mRNA from a variety of lymphoid and nonlymphoid cells revealed multiple mRNA components ranging in size from about 4 to 9.5 kb (Fig. 5). All five of these discrete components, designated *a*–*e*, were detected by a full-length κY3 probe (Fig. 5*A*), the 654-bp *Eco*RI–*Eco*RV fragment (Fig. 5*B*), and a fragment that encodes the N-terminal region of the protein (data not shown). Interestingly, the overall abundance, as well as the relative amounts of individual components, varied significantly among the different cell types. The abundance was generally higher in cells representing early stages of the B lymphoid lineage (i.e., pre-B and B cells) than in cells representing mature plasmacytes or other cell lineages (e.g., fibroblasts). Moreover, component *a* was observed only in

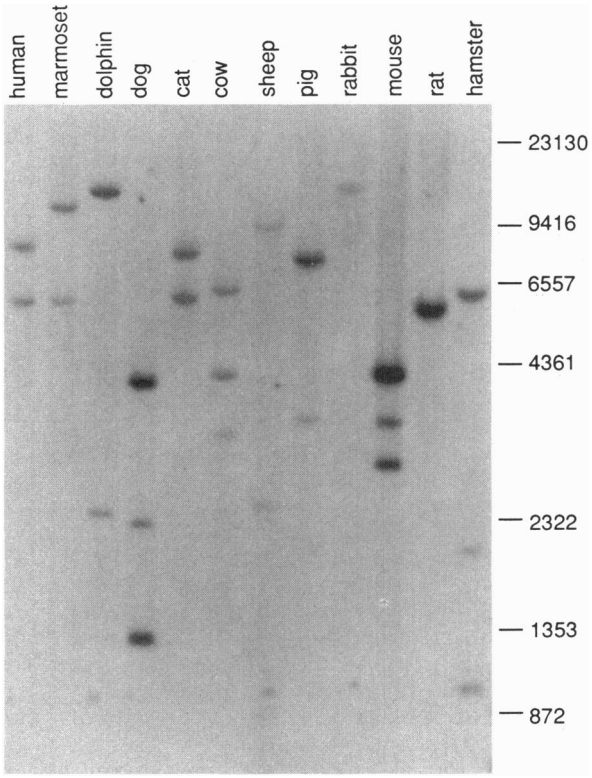

FIG. 4. Distribution of the CHD-1 gene among various mammalian species as determined by Southern blot analysis. *Eco*RI digests of DNA from various mammalian sources were probed with the *Eco*RI–*Eco*RV fragment (see Fig. 1) that spans a portion of the helicase domain. Numbers at right indicate sizes (in base pairs) of λ phage and φX174 RF DNA fragments used as standards.

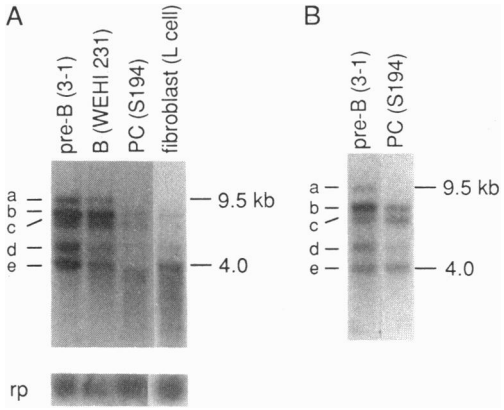

FIG. 5. Differential expression of CHD mRNA among various murine cell lines. Cytoplasmic poly(A)$^+$ mRNA (5 μg) was probed with labeled κY3 insert (*A*) or the *Eco*RI–*Eco*RV fragment (*B*). Five discrete components, *a*–*e*, are indicated. Approximate sizes of components *a* and *e* are indicated at right. Blot in *A* was rehybridized with a mouse ribosomal protein L32 cDNA probe (rp) to control for the amount of RNA loaded on each lane. PC, plasmacytoma.

the early B lymphoid stages, and the relative amounts of components *b* and *d* were significantly lower in the late-stage cells. Some of these multiple mRNA components might be the products of alternative RNA processing and could conceivably encode variant forms of CHD-1 with stage- or lineage-specific roles in gene expression.

DISCUSSION

The discovery of the protein CHD-1, which contains diagnostic features of both the SNF2 (Brm) helicase family and the Pc/HP1-like chromodomain family of proteins, strongly supports the idea that these families are functionally related (3, 4). This relationship has previously been inferred from genetic studies in yeast and *Drosophila*. In yeast, deleterious mutations in the *SNF2/SWI2* gene are suppressed by mutations in genes such as *SPT11*, *SPT12*, *SIN2*, and *SIN1*, which encode histones and HMG1-like proteins. In *Drosophila*, mutations in *Brm*, the counterpart of *SNF2/SWI2*, suppress mutations in *Pc*, which encodes a chromatin-associated protein. These observations have led to the formulation of models in which the putative helicase activity of the SNF2 (Brm) family of proteins plays a role in chromatin remodeling and in counteracting the repressive effects of histones and other proteins involved in gene packaging, such as Pc, HP1, and SIN1 (3, 19). Although the genetic evidence did not demonstrate a direct physical interaction between these two families of proteins, the existence of a protein like CHD-1, which belongs to both families, suggests that such direct interactions do, in fact, occur.

A feature that appears to set CHD-1 apart from other members of the SNF2 (Brm) family is its DNA-binding capability. The DNA-binding domain of CHD-1 was localized to a region that bears no apparent similarity to other proteins in the family. Two segments within this region (amino acid residues 1104–1114 and 1330–1348) have a particularly high density of positively charged amino acids, which could contribute to this property. It is noteworthy that the positively charged segment at 1104–1114 contains the motifs Lys-Arg-Pro-Lys-Lys and Arg-Gly-Arg-Pro-Arg, which enable proteins such as HMG I(Y), D1, and Engrailed to bind in the minor-groove of A+T-rich DNA (see ref. 20 for review).

The sequence specificity of the recombinant fusion protein is not exactly the same as that of the nuclear protein in lymphoid cell nuclear extracts. The κY binding site was initially defined as CTTCCTTA on the basis of DNase I and methylation interference footprints of nuclear extract binding complexes (11). In agreement with this designation, a 4-bp mutation in the center of this site abolished binding of the nuclear protein and reduced the transient expression of transfected κ genes by about 75% (11). In contrast, a probe bearing this same mutation (m4) still binds reasonably well to the recombinant protein. Interestingly, the overall A+T content of the probes that bound strongly to CHD-1 tends to be greater than that of the more weakly binding probes (Table 1). Whether the sequence selectivity of CHD-1 is based on a preference for A+T-rich sequences or on more specific sequence discrimination remains to be established.

The difference in the DNA-binding specificity of CHD-1 and the nuclear extract protein, the large difference in size (197 kDa vs. an estimate of 60–80 kDa, based on the electrophoretic mobility of the nuclear protein–DNA complex relative to that of Oct-1 or Oct-2 complexes), and the presence of detectable amounts of CHD-1 mRNA in nonlymphoid cells would seem to indicate that CHD-1 is not the lymphoid-specific transcription factor that normally interacts with the κY element. Nonetheless, its striking relationship to the SNF2 and chromodomain families of proteins, its sequence-selective DNA-binding capability, and its widespread occurrence among mammals indicate that it has an important role in gene regulation. Moreover, the qualitative and quantitative differences in CHD mRNA components in various cell types suggest that it may exist in different isoforms with distinct functions.

We are deeply indebted to Heide Wilcox and Kenneth Tartof for the Southern blot analysis, to Dawn E. Kelley for supplying some of the RNA samples, to Anita Cywinski for help with DNA sequencing, and to Alfonso Bellacosa and Diane Ruezinsky for help with technical procedures. This research was supported by Grants AI17330, CA06927, and RR05539 from the National Institutes of Health and by an appropriation from the Commonwealth of Pennsylvania. V.D. gratefully acknowledges support from the French Centre National de la Recherche Scientifique and the Association pour la Recherche sur le Cancer.

1. Mitchell, P. J. & Tjian, R. (1989) *Science* **245**, 371–378.
2. Felsenfeld, G. (1992) *Nature (London)* **355**, 219–224.
3. Winston, F. & Carlson, M. (1992) *Trends Genet.* **8**, 387–391.
4. Tamkun, J. W., Deuring, R., Scott, M. P., Kissinger, M., Pattatucci, A. M., Kaufman, T. C. & Kennison, J. A. (1992) *Cell* **68**, 561–572.
5. Messner, S., Franke, A. & Paro, R. (1992) *Genes Dev.* **6**, 1241–1254.
6. Laurent, B. C., Yang, X. & Carlson, M. (1992) *Mol. Cell. Biol.* **12**, 1893–1902.
7. Vinson, C. R., La Marco, K. L., Johnson, P. F., Landschultz, W. H. & McKnight, S. L. (1988) *Genes Dev.* **2**, 801–906.
8. Beckman, H., Su, L.-K. & Kadesch, T. (1990) *Genes Dev.* **4**, 167–179.
9. Gallagher, S. R. & Smith, J. A. (1991) in *Current Protocols in Molecular Biology*, eds. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (Greene and Wiley–Interscience, New York), Unit 10.2.
10. Kelley, D. E., Pollok, B. A., Atchison, M. L. & Perry, R. P. (1988) *Mol. Cell. Biol.* **8**, 930–937.
11. Atchison, M. L., Delmas, V. & Perry, R. P. (1990) *EMBO J.* **9**, 3109–3117.
12. Laurent, B. C., Treitel, M. A. & Carlson, M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 2687–2691.
13. Kozak, M. (1987) *Nucleic Acids Res.* **15**, 8125–8148.
14. Okabe, I., Bailey, L. C., Attree, O., Srinivasan, S., Perkel, J. M., Laurent, B. C., Carlson, M., Nelson, D. L. & Nussbaum, R. (1992) *Nucleic Acids Res.* **20**, 4649–4655.
15. Paro, R. & Hogness, D. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 263–267.
16. Singh, P. B., Miller, J. R., Pearce, J., Kothary, R., Burton, R. D., Paro, R., James, T. C. & Gaunt, S. J. (1991) *Nucleic Acids Res.* **19**, 789–794.
17. Pearce, J. J. H., Singh, P. B. & Gaunt, S. J. (1992) *Development* **114**, 921–929.
18. Riggs, P. (1992) in *Current Protocols in Molecular Biology*, eds. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (Greene and Wiley–Interscience, New York), Unit 16.6.
19. Travers, A. A. (1992) *Cell* **69**, 573–575.
20. Churchill, M. E. A. & Travers, A. A. (1991) *Trends Biochem. Sci.* **16**, 92–97.
21. James, T. C. & Elgin, S. C. R. (1986) *Mol. Cell. Biol.* **6**, 3862–3872.