Molecular Cloning of the α -Globin Transcription Factor CP2

LORENA C. LIM, STEVEN L. SWENDEMAN, AND MICHAEL SHEFFERY*

*DeWitt Wallace Research Laboratory and the Graduate Program in Molecular Biology,
Memorial Sloan-Kettering Cancer Center, New York, New York 10021*

Received 26 August 1991/Accepted 11 November 1991

CP2, a transcription factor that binds the murine α -globin promoter, was purified and subjected to amino acid sequence analysis. Oligonucleotide primers derived from the sequence were used to obtain murine and human cDNA clones for the factor. The murine cDNA spans approximately 4 kb and contains two coextensive open reading frames (ORFs) which encode deduced polypeptides of 529 (ORF-1; molecular weight, 59,802) and 502 (ORF-2; molecular weight, 56,957) amino acids, slightly smaller than the purified factor as estimated from its mobility in sodium dodecyl sulfate-polyacrylamide gels (64,000 to 66,000). The human cDNA contains a single ORF of 501 amino acids that is nearly contiguous with murine ORF-2. Indeed, comparison of deduced human and murine amino acid sequences shows that the two polypeptides are 96.4% identical. A strictly conserved region is rich in serine and threonine (17.5%) and in proline (11%) residues (S-T-P domain). This S-T-P domain is immediately amino terminal to a string of 10 glutamines (in the human sequence) or a tract of alternating glutamine and proline residues (in the mouse sequence). Bacterial expression of the full-length (502-amino-acid) murine factor or of a core region comprising amino acids 133 to 395 generated polypeptides with the DNA binding specificity of CP2. These results confirmed the cloning of CP2 and delimited the region sufficient for specific DNA sequence recognition. Antisera produced against the core region recognized polypeptide species with M_r s of 64,000 and 66,000 in immune blots of nuclear extracts prepared from both murine and human cell lines, consistent with the size of the purified factor. Lastly, a data base search revealed that amino acids 63 to 270 of the murine factor are distantly related to a domain in the *Drosophila* gene regulatory factor Elf-1.

We have previously described the purification and characterization of several α -globin promoter transcription factors present in nuclear extracts prepared from murine erythroleukemia (MEL) cells (1, 2, 9-11). These factors include the tissue-restricted transcription factor GATA-1, the heterotypic CCAAT factor CP1, and a third factor termed CP2. Purified CP2 (a polypeptide doublet with M_r s of 64,000 and 66,000 on sodium dodecyl sulfate [SDS]-polyacrylamide gels [9]) interacts with a region of the α -globin promoter that extends from -85 (where it overlaps with the binding site of the α -globin CCAAT factor, CP1) to -160, just downstream from the GATA-1 binding site (1, 11). The factor binds a consensus sequence (GCN[C/A]NANCCAG) repeated four times within this region (11). Factor binding activity is observed in a wide variety of cell types, including the human cell lines HeLa (4) and CEM. Our previous results have shown that CP2 binding activity is modestly induced (two- to threefold) during erythroid differentiation of MEL cells (1). Additionally, in vitro transcription experiments have shown that CP2 is present at functionally limiting amounts in MEL cell nuclear extracts (11) and that addition of DNA sequence affinity-purified factor stimulates a 10-fold increase in the transcription of templates constituted from a single CP2 factor binding site juxtaposed to the α -globin TATAA element (11). To improve our understanding of the structure and biosynthesis of CP2 and to begin to dissect how the factor helps activate α -globin gene transcription, we have obtained murine and human cDNA clones that encode CP2.

MATERIALS AND METHODS

Cell lines and tissue culture. MEL cells were maintained in roller or spinner bottles in alpha minimal essential medium (GIBCO) supplemented with 5% enriched calf serum (Gemini Bioproducts) at 37°C in 5% CO₂. Cells were induced by culture in the presence of 5 mM hexamethylenedisacetamide (Aldrich). HeLa and CEM cells were maintained in RPMI 1640 (GIBCO) supplemented with 10% fetal calf serum.

Nuclear extracts. Nuclear extracts were prepared as described previously (6, 11) except that all buffers contained leupeptin (1 μ g/ml), pepstatin A (1 μ g/ml), aprotinin (1 μ g/ml), phenylmethylsulfonyl fluoride (PMSF; 1 mM), benzamide (2 mM), and NaF (10 mM).

Purification of CP2 and amino acid sequence determination. CP2 was purified from induced MEL cell extracts as described previously (9). Briefly, nuclear extracts were applied sequentially to phosphocellulose, DEAE, single-stranded DNA-cellulose, and a CP2 DNA sequence affinity column. CP2 DNA binding activity was monitored during chromatography by electrophoretic gel shift assays as described previously (7, 20). Approximately 20 μ g of CP2 at a purity of 90% was obtained from 2×10^{10} cells. Purified protein was precipitated with 4 volumes of acetone, applied to an SDS-polyacrylamide gel (13), and transferred to nitrocellulose as described previously (15). CP2 was visualized with 0.1% Ponceau S in 1% acetic acid. Proteolytic digestion with trypsin and tryptic peptide separation were performed as described previously (24). Selected peptides were sequenced on an Applied Biosystems 477A sequencer.

Preparation of CP2 cDNA probes by PCR. Sense and antisense guessmers deduced from tryptic peptide sequences were synthesized and used to generate a polymerase chain reaction (PCR) clone of an internal region of CP2 as described previously (23). Briefly, 1 μ g of poly(A)⁺ RNA from

* Corresponding author.

induced MEL cells was incubated in 25 μ l with 50 pM appropriate antisense primer and 50 U of Moloney reverse transcriptase (Bethesda Research Laboratories) in 50 mM Tris (pH 8.3)–75 mM KCl–3 mM MgCl₂–10 mM dithiothreitol–200 mM deoxynucleoside triphosphates (dNTPs)–25 U of RNasin (Promega) for 1 h at 37°C. Samples were extracted with phenol, and cDNA was recovered by ethanol precipitation. cDNAs were amplified in 100- μ l reaction mixtures containing 10 mM Tris (pH 8.3), 50 mM KCl, 2.5 mM MgCl₂, 200 μ M dNTPs, 50 pM each appropriate sense and antisense primers, and 2.5 U of *Taq* DNA polymerase (Perkin-Elmer/Cetus). Amplified fragments were subcloned into the pCR 1000, TA cloning vector (Invitrogen), and the DNA sequence was obtained by the dideoxy method (19). One clone of 817 bases had a single open reading frame (ORF) that encoded each of the tryptic peptides. This sequence was purified and used to screen MEL and HeLa cell cDNA libraries prepared in lambda ZAP (Stratagene) as described previously (18). Five clones were isolated; inserts were excised (zapped) from the phage vector into Bluescript plasmids and sequenced (19).

Northern (RNA) and Southern blot analysis. Northern and Southern blots were prepared, hybridized, washed, and visualized by autoradiography as described previously (18).

Electrophoretic gel shift assays. Electrophoretic shift assays were performed as described previously (7, 20), using a kinase-treated, double-stranded oligonucleotide corresponding to the α -globin CP2 DNA recognition sequence: 5'-GATCCCAAGTTTACTCGGTAGAGCAAGCACAAACCAGG-3'/5'-GATCCCTGGTTTGTGCTTGCTCTACCGAGTAAACTTGG-3' (11). A double-stranded oligonucleotide containing a CP1 binding site and used as nonspecific competitor for CP2 binding activity has the sequence 5'-AGCTCAAATTAACCAATCAGCGCACTCTCACAGG-3'/5'-AGCTCCTGTGAGAGTGCCTGATTGGTTAATTG-3' (10).

Gel electrophoresis and immunoblot analysis. SDS-polyacrylamide gel electrophoresis (PAGE) was performed as described previously (13). For visualization of bacterially synthesized protein, gels were stained with 0.1% Coomassie brilliant blue in 50% methanol–10% acetic acid.

Immunoblotting was performed as described previously (25). Briefly, after SDS-PAGE, samples were transferred by electroblotting to nitrocellulose in 160 mM glycine–2.5 mM Tris–20% methanol. Filters were incubated for 1 h in phosphate-buffered saline (PBS) supplemented with 5% nonfat dry milk, then washed once in PBS to eliminate excess milk and incubated with either preimmune or immune serum diluted in incubation buffer (PBS containing 300 mM NaCl and 0.1% Tween 20 [Bio-Rad]). Final serum dilutions were 1:300. After 2 to 4 h in the presence of the appropriate serum, blots were washed at room temperature for 30 min in 250 ml of incubation buffer and then exposed for 30 min to incubation buffer containing an alkaline phosphatase conjugate of goat anti-rabbit immunoglobulin G (1:1,000 dilution; Sigma). Blots were washed as described above, and immune reacting species were visualized by incubation in 50 mM Tris–50 mM NaCl–10 mM MgCl₂ containing 1 mg of nitro-blue tetrazolium (Sigma) per ml and 0.5 mg of 5-bromo-4-chloro-3-indolyl phosphate (Sigma) per ml for 5 to 15 min.

Preparation of immunogens and bacterially expressed CP2. Regions of the murine CP2 cDNA corresponding to (from ORF-2) amino acids 1 to 502 or 133 to 395 were amplified by PCR techniques (adding sequences to encode an initiator methionine in the case of the internal polypeptide sequence) and expressed by using the pET-3c vector system (14, 17). Induced protein was enriched as described previously (11).

FIG. 1. Northern and Southern blot analyses of CP2. (A) An 817-base clone obtained from the PCR amplification of MEL cell mRNA was used to probe poly(A)⁺ mRNA prepared from induced MEL cells at 60 (lane 1) or 100 (lane 2) h after induction. Mobilities of 28 and 18S rRNAs are shown at the left. (B) The same probe was used to analyze DNA prepared from MEL cells (mouse) and the human T-cell line CEM (human). DNA was restricted with either *Eco*RI (lane E), *Pst*I (lane P), or *Bgl*II (lane B). Mobilities of DNA size markers are indicated at the right in kilodaltons.

Briefly, after induction in the presence of 0.5 mM isopropylthiogalactoside (IPTG), bacterially expressed CP2 was isolated by harvesting cells ($7,000 \times g$ for 10 min) and suspending them in 20 mM Tris (pH 8.0)–10 mM EDTA–50 mM glucose containing 2 mg of lysozyme per ml. After 30 min at 4°C, cells were pelleted ($10,000 \times g$ for 10 min), suspended in 10 mM Tris (pH 8.0)–1 mM EDTA–10% glycerol–1 M NaCl–4 M urea–1 mM PMSF, and incubated at 70°C for 5 min. The supernatant was recovered after centrifugation ($10,000 \times g$ for 10 min) and was dialyzed for 3 h against 4 liters of 20 mM Tris (pH 7.9)–20% glycerol–100 mM NaCl–0.2 mM EDTA supplemented with dithiothreitol (0.5 mM), leupeptin (1 μ g/ml), pepstatin A (1 μ g/ml), aprotinin (1 μ g/ml), PMSF (1 mM), benzamide (2 mM), and NaF (10 mM). The dialysate was clarified by centrifugation ($16,000 \times g$) and used for subsequent electrophoretic gel shift experiments or for preparative SDS-polyacrylamide gels. In the latter case, gel regions corresponding to the expressed portion of CP2 were identified by staining with Coomassie brilliant blue, excised from the gel, crushed with a glass rod, and extracted in 0.1 M NaHCO₃ buffer (pH 8.3) supplemented with 0.1% SDS and the protease and phosphatase inhibitors described above. Protein eluted in this manner was used to prepare rabbit antisera (see below).

Rabbit antisera. The bacterially expressed core portion of CP2, purified from SDS-polyacrylamide gels, was emulsified in Freund's complete adjuvant (GIBCO) and administered subcutaneously to New Zealand White rabbits (100 to 150 μ g per injection). Rabbits were boosted at days 25, 35, and 45 with immunogen (100 to 150 μ g) emulsified in Freund's incomplete adjuvant. Blood was collected from dorsal ear veins and allowed to clot overnight at 4°C. Sera were collected and clarified by centrifugation, and samples were

830

FIG. 2. Nucleic acid and deduced protein sequences of human and murine CP2. (A) Nucleic acid sequence (numbered to the left) and the deduced amino acid sequence (right) of a murine cDNA encoding CP2. Amino acids corresponding to sequenced peptides are underlined. Note that the murine cDNA has two coextensive ORFs, only one of which is shown. Inclusion of the upstream amino acids would add the 27-amino-acid string MFRLRRVWGRLLGGTLGAGACWGKEGAG to the amino terminus of the sequence shown. (B) Nucleic acid sequence and deduced amino acid sequence of the human cDNA clone encoding CP2. Markings are as in panel A.

stored at -70°C . Preimmune sera were collected and stored in a similar manner.

Nucleotide sequence accession number. GenBank accession numbers for the murine and human CP2 cDNA sequences are M76987 and M76988, respectively.

RESULTS

Cloning of CP2. Approximately 30 μg of affinity-purified CP2 was separated on an SDS-polyacrylamide gel and transferred to nitrocellulose. Protein was visualized with Ponceau S, and the two polypeptides that compose CP2 binding

MOUSE	MAWALKLPLADEVIESGLVQDFDASLSGGIGQELGAGAYSMSDVLALPIFKQEESLPPDM	60
		
HUMAN	MAWALKLPLADEVIESGLVQDFDASLSGGIGQELGAGAYSMSDVLALPIFKQEESLPPDM	60
MOUSE	ENEILFPQYVLC AATSFAVKLEDETTLTYLNQAGSYEIRMLDMRKLGE LPELNGKLVKSIF	120
	!!	
HUMAN	ENKILFPQYVLC AATSFAVKLEDETTLTYLNQAGSYEIRMLDMRKLGE LPEINGKLVKSIF	120
MOUSE	RVVFHDERRLQYTEHQQLLEGWRWNRPGDRILDIDIPMSVGVIDPRANPTQLNTVEFLWDPS	180
		
HUMAN	RVVFHDERRLQYTEHQQLLEGWRWNRPGDRILDIDIPMSVGIIIDPRANPTQLNTVEFLWDPA	180
MOUSE	KRTSVFIQVHCISTEFTMRKHGGEKGVVFRVQIDTFKENGNGEYTHELHSASCQIKVFKP	240
		
HUMAN	KRTSVFIQVHCISTEFTMRKHGGEKGVVFRVQIDTFKENENG EYTHELHSASCQIKVFKP	240
MOUSE	KGADRKQKIDREKMEKRTPEKEKYQPSYETITLTCSPWPEITYVMNSPSPGFNSSHS	300
		
HUMAN	KGADRKQKMDREKMEKRTPEKEKYQPSYETITLTCSPWPEITYVMNSPSPGFNSSHS	300
MOUSE	FSLGEGNGSPMHQPEPPPPFVTDMLLPTTTPQEAQQWLHRMRFS TTRLFNFSGADLLKL	360
		
HUMAN	FSLGEGNGSPMHQPEPPPPFVTDMLLPTTTPQEAQQWLHRMRFS TTRLFNFSGADLLKL	360
MOUSE	TRDDVIQICGPADGIRLFNALKGRMVRPRLTIYVCQESLQLREQQPQPQKQEDGDS	420
		
HUMAN	TRDDVIQICGPADGIRLFNALKGRMVRPRLTIYVCQESLQLREQQQQQQQQKQEDGDS	420
MOUSE	NGTFFVYHAIYLEELTAVELTEKIAQLFSISPEQISQIYKQGPTG IHVVISDENVNQNFQE	480
		
HUMAN	NGTFFVYHAIYLEELTAVELTEKIAQLFSISPCQISQIYKQGPTG IHLVISDENIQNFQE	480
MOUSE	EACFILDTEAETS DSYHVLK	502
		
HUMAN	EACFILDTEK-ETNDSYHILK	501

FIG. 3. Comparison of deduced mouse and human amino acid sequences. Amino acid numbers are indicated at the right. Only identical amino acids are indicated (by double dots). The two sequences are 96.4% identical, and most amino acid differences represent conservative changes.

activity (M_r s of 64,000 and 66,000) were excised separately. Because the M_r -64,000 polypeptide is more abundant, we proceeded to characterize this species.

Attempts to obtain amino-terminal sequence data from purified CP2 were unsuccessful. Therefore, nitrocellulose strips containing CP2 were subjected to digestion with trypsin. Tryptic peptides were separated by high-performance liquid chromatography and collected for amino acid sequence analysis. The sequences of the four peptides were as follows: ANPTQLNTVEFL (peptide 1); LQYTEHQQLLEGW (peptide 2); LFTNFSGADLL (peptide 3); and LTIYVCQE (peptide 4). Sense and antisense oligonucleotide guessmers corresponding to each peptide were synthesized, and antisense oligonucleotides were used to prime reverse transcription of poly(A)⁺ RNA prepared from induced MEL cells. Potential cDNAs were amplified by PCR using appropriate primer pairs. One primer pair (designed to amplify sequences between peptides 1 and 4) generated a DNA fragment of approximately 700 bases, whereas a second pair (appropriate to amplify sequences between peptides 2 and 4) generated a doublet of slightly larger fragments (approximately 800 bases; not shown). The latter PCR products were cloned and sequenced. Inspection of the DNA sequences revealed that one PCR product, of 817 bases, contained a single ORF. As expected, this ORF began with peptide 2 and terminated with peptide 4. Importantly, peptides 1 and 3 were also identified within the ORF, and the DNA sequence encoding the former peptide was found approximately 100 bases into the clone, consistent with the PCR results described above.

These results suggested that the cloned PCR product contained a portion of the CP2 cDNA. To estimate the size of the corresponding full-length message, the 800-base se-

quence was used to probe Northern blots of poly(A)⁺ RNA prepared from MEL cells. The probe detected a single mRNA species of approximately 3,800 bases (Fig. 1A). In addition, when used to analyze Southern blots of murine and human DNAs, the probe recognized a limited number of restriction fragments in both species (Fig. 1B). This latter result, together with the fact that CP2 binding activity is readily detected in nuclear extracts prepared from HeLa cells, suggested that the cloned murine probe could be used to obtain human as well as murine cDNAs.

Characterization of the murine and human cDNAs. By using the probe characterized above, cDNA libraries prepared from MEL and HeLa cells were screened, and several clones were isolated and sequenced. Figure 2 shows the nucleotide sequences of the largest murine and human cDNAs obtained, and Fig. 3 directly compares their deduced amino acid sequences. The murine clone extends 4,023 bases and includes an 871-base 5' leader, a 1,506-base ORF, and 1,644 bases 3' of the putative termination codon. The size of the clone is consistent with the estimate of CP2 mRNA obtained by Northern blot analysis (Fig. 1) and suggests that it is close to full length. The largest human cDNA clone obtained was 2.4 kb (Fig. 1B). This clone lacked a poly(A) tail, presumably because the HeLa cDNA library was prepared by using random primers in the reverse transcription reaction.

Analysis of the murine cDNA revealed two coextensive ORFs that encode deduced polypeptides of 529 (ORF-1) and 502 (ORF-2) amino acids. The calculated M_r s of these polypeptide sequences are 59,802 and 56,957, respectively, slightly smaller than the observed M_r of the native factor estimated from SDS-polyacrylamide gels (9). Although there

FIG. 4. Sequence-specific DNA binding of the cloned factor. (A) Bacterial expression of murine ORF-2. A Coomassie-stained 10% SDS-polyacrylamide gel is shown. Lanes: 1, markers with indicated M_r s (in thousands); 2 and 3, uninduced (lane 2) and induced (lane 3) bacterial cultures; 4 and 5, urea extractions of two duplicate cultures (see Materials and Methods); 6, urea extraction of uninduced culture. (B) Electrophoretic gel shifts generated by bacterially expressed CP2. Lanes: 1, CP2 probe alone (1 ng); 2, induced bacterial extract (2 μ l); 3, uninduced bacterial extract (2 μ l); 4 to 10, equal amounts (2 μ l) of induced bacterial extract and no competitor (lane 4), 100, 200, and 300 ng of unlabeled probe DNA (lanes 5 to 7, respectively), and 100, 200, and 300 ng of an oligonucleotide specific for the CCAAT factor, CP1 (lanes 8 to 10, respectively). (C) Bacterial expression of the core region of CP2. A Coomassie-stained 12.5% SDS-polyacrylamide gel is shown. Lanes: 1, markers with M_r s of 68,000 and 43,000; 2 to 5, samples from duplicate uninduced (lanes 2 and 3) or induced (lanes 4 and 5) bacteria. The urea extraction of induced species is comparable in purity to that illustrated in panel A (not shown). (D) Electrophoretic gel shifts of a core region of CP2. Lanes: 1, CP2 probe alone (1 ng); 2, induced bacterial extract (12 μ l); 3, uninduced bacterial extract (12 μ l); 4 to 8, equal amounts (12 μ l) of induced bacterial extract and no competitor (lane 4), 100 and 200 ng of unlabeled probe DNA (lanes 5 and 6, respectively), and 100 and 200 ng of oligonucleotide specific for the CCAAT factor CP1 (lanes 7 and 8, respectively).

are two contiguous ORFs in the murine cDNA, the translation initiation region at ORF-2 has a better match to a consensus initiation sequence (12) than does ORF-1 (9 of 13 versus 5 of 13). In addition, comparison with the derived amino acid sequence of the human factor shows that the two polypeptides are nearly identical when compared from murine ORF-2 but diverge if translation is continued further upstream, to the beginning of ORF-1. For these reasons, we have illustrated the murine sequence beginning at ORF-2.

Comparison of the mouse and human polypeptides (Fig. 3) showed that the two sequences are 96.4% identical. Indeed, a region of complete sequence identity spans 154 amino acids (amino acids 250 to 405). This region is rich in serine and threonine (17.5%) and proline (11%) residues (S-T-P domain). In addition, this S-T-P domain contains several SPXX motifs, an amino acid sequence observed in many DNA-binding proteins and transcription factors (21, 22). The

FIG. 5. Immunoblot analysis. Nuclear extracts prepared from MEL cells (lanes 1, 2, 4, and 6) or the human T-cell line CEM (lanes 3 and 5) were separated on a 10% SDS-polyacrylamide gel, transferred to nitrocellulose, and incubated with preimmune (lanes 1 to 3) or immune (lanes 4 to 6) sera. Approximately 50 μ g of protein was applied to each lane except lane 6 (20 μ g). Immune sera were against CP2 (lanes 4 and 5) or GATA-1 (lane 6). CEM cells contain no species that react with the GATA-1 serum (not shown). The mobilities of species that react with the CP2 and GATA-1 antisera are indicated at the right, and those of protein standards are indicated at the left (numbers indicate M_r s [in thousands]).

region of strict conservation is preceded by a slightly basic region and is followed by a stretch of 10 glutamines in the human protein or a sequence of alternating glutamine and proline residues (Q/P domain) in the mouse. This latter motif is also reminiscent of DNA binding and transcriptional activation domains described previously (5, 16). The deduced amino acid sequence distal to the Q/P domain has a moderate net negative charge (−10).

Further comparison showed that most of the differences observed between the two polypeptides represent conservative exchanges. Three exceptions were noted. Twice, a glutamic acid residue in the mouse sequence is replaced by lysine in the human polypeptide (at murine residues 63 and 490). While this exchange dramatically alters local charge, it would conserve a hydrophilic character. The third exception occurs in the short (10-residue) polyglutamine tract seen in the human sequence. In the mouse, these residues appear as alternating glutamine and proline residues. This clustering of substitutions (in two such similar polypeptides) suggests that the proline and glutamine residues found in this region are functionally equivalent.

Bacterial expression of CP2. To confirm that the cloned sequence encodes CP2, we inserted murine ORF-2 into an inducible prokaryotic expression vector (14, 17) by using PCR methods. Visualization of bacterial lysates prepared before or after induction on Coomassie-stained SDS-polyacrylamide gels revealed the induced accumulation of a polypeptide species with an M_r of 63,000, consistent with the size of the purified factor (Fig. 4A) (9). Induced bacterial lysates were tested for the presence of CP2 DNA binding activity by electrophoretic gel shift assays. Figure 4B shows that the bacterial protein bound a target sequence containing a CP2 binding site. (We note that the shift generated by the bacterial factor is identical to that seen in nuclear extracts prepared from MEL cells [not shown].) As expected, the shift was abolished by addition of an unlabeled oligonucleotide competitor containing the cognate recognition sequence

Our previous results have demonstrated that CP2 is a strong activator of α -globin transcription in vitro (11). In this regard, we note that a tract of 10 glutamines in the human protein and a sequence of alternating glutamine and proline residues in the mouse cDNA lie just downstream from the conserved S-T-P region. In addition, the carboxy-terminal end of the molecule has a net charge of -10 . Comparison with known transcription factors (5, 8, 16) suggests that these domains might constitute a part of the transcriptional activating region of CP2. Experiments to test these hypotheses are under way.

We have previously shown that CP2 DNA binding activity is modestly induced during erythroid differentiation of MEL cells (1). The probe DNAs and antisera now available should help us define unequivocally the mechanisms that account for the fluctuations of CP2 DNA binding activity during erythroid differentiation. These reagents will also help to define what influence, if any, the increased level of CP2 DNA binding activity exerts on the large scale induction of α -globin gene transcription observed during terminal erythroid differentiation.

ACKNOWLEDGMENTS

This work was supported in part by Public Health Service grants CA-31768 and CA-08748 from the National Cancer Institute, by grants DK-37513 and DK-41079 from the NIH, and by the Businessmen's Assurance Fund.

The technical assistance of Elizabeth Wick is gratefully acknowledged, as is the excellent assistance provided by Scott Geromanos of the Sloan-Kettering Institute microchemistry facility.

REFERENCES

1. Barnhart, K. M., C. G. Kim, S. S. Banerji, and M. Sheffery. 1988. Identification and characterization of multiple erythroid cell proteins that interact with the promoter of the murine α -globin gene. *Mol. Cell. Biol.* 8:3215-3226.
2. Barnhart, K. M., C. G. Kim, and M. Sheffery. 1989. Purification and characterization of an erythroid cell-specific factor that binds the murine α - and β -globin genes. *Mol. Cell. Biol.* 9:2606-2614.
3. Bray, S. J., B. Burke, N. H. Brown, and J. Hirsh. 1989. Embryonic expression pattern of a family of Drosophila proteins that interact with a central nervous system regulatory element. *Genes Dev.* 3:1130-1145.
4. Chodosh, L. A., A. S. Baldwin, R. W. Carthew, and P. A. Sharp. 1988. Human CCAAT-binding proteins have heterologous subunits. *Cell* 53:11-24.
5. Courey, A. J., D. A. Holtzman, S. P. Jackson, and R. Tjian. 1989. Synergistic activation by the glutamine-rich domains of human transcription factor Sp1. *Cell* 59:827-836.
6. Dignam, J. D., P. L. Martin, B. S. Shastri, and R. G. Roeder. 1983. Eukaryotic gene transcription with purified components. *Methods Enzymol.* 101:582-598.
7. Garner, M. M., and A. Rezin. 1981. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. *Nucleic Acids Res.* 9:3047-3060.
8. Giniger, E., and M. Ptashne. 1987. Transcription in yeast activated by a putative amphipathic alpha helix linked to a DNA binding unit. *Nature (London)* 330:670-672.
9. Kim, C. G., K. M. Barnhart, and M. Sheffery. 1988. Purification of multiple erythroid cell proteins that bind the promoter of the α -globin gene. *Mol. Cell. Biol.* 8:4270-4281.
10. Kim, C. G., and M. Sheffery. 1990. Physical characterization of the purified CCAAT transcription factor, α -CP1. *J. Biol. Chem.* 265:13362-13369.
11. Kim, C. G., S. L. Swendeman, K. M. Barnhart, and M. Sheffery. 1990. Promoter elements and erythroid cell nuclear factors that regulate α -globin gene transcription in vitro. *Mol. Cell. Biol.* 10:5958-5966.
12. Kozak, M. 1989. The scanning model for translation: an update. *J. Cell Biol.* 108:229-241.
13. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* 227:680-685.
14. Lin, S.-W., J. J. Dunn, F. W. Studier, and D. W. Stafford. 1987. Expression of human factor IX and its subfragments in Escherichia coli and generation of antibodies from the subfragments. *Biochemistry* 26:5267-5274.
15. Matsudaira, P. 1987. Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *J. Biol. Chem.* 262:10035-10038.
16. Mitchell, P. J., and R. Tjian. 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* 245:371-378.
17. Rosenberg, A. H., B. N. Lade, D.-S. Chui, S.-W. Lin, J. J. Dunn, and F. W. Studier. 1987. Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene* 56:125-135.
18. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
19. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463-5467.
20. Strauss, F., and A. Varshavsky. 1984. A protein binds to a satellite DNA repeat at three specific sites that would be brought into mutual proximity by DNA folding in the nucleosome. *Cell* 37:889-901.
21. Suzuki, M. 1989. SPXX, a frequent sequence motif in gene regulatory proteins. *J. Mol. Biol.* 207:61-84.
22. Suzuki, M. 1989. SPKK, a new nucleic acid-binding unit of protein found in histone. *EMBO J.* 8:797-804.
23. Tan, J. C., N. Nocka, P. Ray, P. Traktman, and P. Besmer. 1990. The dominant W⁴² spotting phenotype results from a missense mutation in the c-kit receptor kinase. *Science* 247:209-212.
24. Tempst, P., A. J. Link, L. R. Riviere, M. Fleming, and C. Elicone. 1990. Internal sequence analysis of proteins separated on polyacrylamide gels at the submicrogram level: improved methods, applications and gene cloning strategies. *Electrophoresis* 11:537-553.
25. Towbin, H., T. Staehelin, and J. Gordon. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. USA* 76:4350-4354.