Nucleic Acids Research, Vol. 19, No. 20 5661–5667

Members of the zinc finger protein gene family sharing a conserved N-terminal module

Margherita Rosati, Maria Marino, Annamaria Franzè, Anna Tramontano[1] and Giovanna Grimaldi*
International Institute of Genetics and Biophysics, Via Marconi 12, Naples 80125 and [1]Istituto Ricerche Biologia Molecolare, Via Pontina Km 30.600, Pomezia, Rome, Italy

Received July 3, 1991; Revised and Accepted September 20, 1991 EMBL accession nos X60152–X60156 (incl.)

ABSTRACT

We report the isolation of human members of a sub-family of structurally related finger protein genes. These potentially encode polypeptides containing finger motifs of the *Krüppel* type at the C-terminus, and a conserved amino acid module at the N-terminus; because of its invariant location the latter is referred to as finger preceding box (FPB). The FPB, detected also in previously described finger proteins from human, mouse and *Xenopus*, extends over approximately 65 amino acids and appears to be composed of two contiguous modules: FPB-A (residues 1–42) and FPB-B (residues 43–65). The latter is absent in some of the members analyzed. Elements A and B and the zinc finger domain are encoded by separate exons in the ZNF2 gene, a human member of this sub-family. The positioning of introns within this gene is remarkable. One intron flanks and a second interrupts the first codon of the FPB-A and FPB-B modules, respectively. A third intron occurs a few nucleotides downstream of FPB-B marking its separation from the remainder of the coding sequences. This organization, together with the absence of FPB-B in some cDNAs, supports the hypothesis that mRNAs encoding polypeptides that include one, both or none of the FPB-A and FPB-B modules may be assembled through alternative splicing pathways. Northern analyses showed that members of this sub-family are expressed as multiple transcripts in several cell lines. The sequences of distinct cDNAs homologous to the ZNF2 gene indicate that alternative splicing events adjoin either coding or non coding exons to the FPB sequences.

INTRODUCTION

Zinc fingers, amino acid stretches featuring regularly spaced pairs of cysteines and histidines or differently arranged cysteines organized around zinc(II) ions (1–3), are considered major structural motifs involved in protein-nucleic acid interaction (4–7). Finger motifs of the C_2H_2 type, originally described as the potential functional units that mediate the binding of the TFIIIA factor (8, 9) to the *Xenopus* 5S RNA gene promoter, have been subsequently found in a variety of known or putative DNA binding regulatory proteins. Transcription factors from different species, such as yeast ADR1 and human Sp1 (10, 11), as well as proteins encoded by genes playing a pivotal role in *Drosophila* development (12, 13) contain arrays of these motifs. The evolutionary conservation of a linker stretch (the H/C link; 14), first identified in the *Krüppel* segmentation gene from *Drosophila* (12), which connects adjacent zinc fingers has been exploited to clone a number of finger-encoding genes from several species by cross-hybridization (14–26). Additional members of this large family were isolated because of their properties. Murine Krox-20 (27) and Krox-24 (28), as well as their human counterparts EGR1 (29) and EGR2 (30), are responsive to early growth stimuli; GLI (31), Evi-1 (32) and XT33 (33) are rearranged in tumor cells.

The sequences of several potential zinc finger polypeptides have been reported to date. Additional common elements outside of the finger repeats were found in cDNA clones from *Xenopus*. In that instance comparative analysis of the potentially encoded ORFs led to the identification of several discrete motifs shared by the products of distinct finger genes (FAX boxes; 34). Conserved modules in regulatory proteins specify properties relevant to their action; the occurrence of similar motifs in distinct genes may therefore represent the first indication of functionally related protein subsets.

Here we report the characterization of human members of a sub-family of zinc finger genes encoding polypeptides that feature a highly conserved module located at their N-terminus.

MATERIALS AND METHODS

Nomenclature of zinc finger recombinant clones

Clones from our laboratory representative of human finger genes which have been assigned a ZNF symbol from the Human Gene Mapping Committee (35) are indicated accordingly; those which do not fall in this category are named with a HZF (human zinc finger) symbol. In either cases the numbers (X.Y) identify different (Y) homologous (X) clones. The origin of clones is indicated (c, cDNA; g, genomic).

* To whom correspondence should be addressed

Library Screenings, Southern Blotting, DNA Sequencing and Sequence Analysis

cZNF2.2 was isolated upon screening of a λgt10 cDNA library from undifferentiated human teratocarcinoma NTera2-D1 cells (36) using a DNA segment derived from clone cZNF2.1 (see legend to Figure 1). The latter was isolated from a cDNA library in λ T7-T3 E-H (37), prepared from human liver poly(A)$^+$ RNA as described by Huynh *et al.* (38), probed with a 500 bp long Eco RI fragment from the murine clone mKr1 (17) under reduced stringency (hybridization in 6×SSC, 10× Denhardt's solution, 0.5% SDS at 58°C; washes in 2×SSC, 0.5% SDS at 50°C). A collection of zinc finger recombinants was obtained upon isolation of cDNA clones from the λgt10 cDNA library mentioned above or an equivalent one in λgt11 (both were a kind gift of J. Skowronski and M.F. Singer), using as probe a 700 bp Eco RI-Hind III fragment from the human finger gene ZNF41 (26) under the conditions described above. Clones cHZF3.1, cZNF41.1 and cHZF7.1 were sorted out from this collection using as a probe a DNA segment from clone cZNF2.2 (probe B, Figure 1A). The conditions used were: hybridization in 6×SSC, 10× Denhardt's solution, 0.5% SDS at 58°C; sequential washes in 6×SSC, 0.5% SDS from room temperature to 45°C. Recombinants including the ZNF2 gene (gZNF2 clones) were isolated from a human genomic library in pcos2EMBL (39) using a DNA segment from cZNF2.2 (probe A, Figure 1A).

All probes were labelled to a specific activity of $2-5\times10^8$ cpm/mg by random-hexanucleotide-priming as previously described (26). DNA preparations, gel electrophoresis and Southern blotting were performed according to standard procedures (40). DNA fragments of interest were subcloned into Blue Script (Stratagene). DNA sequences were determined, by the dideoxy chain-termination method, from double stranded plasmids, phage and cosmid DNAs according to described procedures (41), using the Sequenase kit from United States Biochemical (Cleveland, Ohio).

NBRF protein sequence data base search was performed using the program FASTA of the GCG sequence analysis package (42); the sequence used as a probe corresponded to amino acids 15−79 of the cZNF2.2 ORF. Amino acid sequence similarities were identified by pairwise comparisons using the ISSC (Interactive Sensitive Sequence Comparison; 43) program. The COLLATE package of secondary structure prediction methods (44−46) was kindly provided by Chris Sander (EMBL, Heidelberg).

Cell Cultures, RNA isolation and Northern analysis

The human cell lines used are: NTera2-D1 (teratocarcinoma); HepG2 (hepatoblastoma); HeLa (cervical carcinoma); MOLT-4 (T-cell lymphoma); HL-60 (acute promyelocytic leukemia); Jurkat (B-cell lymphoma); HT-1080 (fibrosarcoma); A1251 (kidney carcinoma). HepG2, HeLa, HT-1080, A1251 and NTera2-D1 cells were cultured in DMEM, 10% fetal calf serum; MOLT-4, HL-60 and Jurkat cells were grown in RPMI 1640, 10% fetal calf serum.

Five μg of poly(A)$^+$ RNA, prepared from total RNA (26) were electrophoresed on 1% agarose-formaldehyde gels and transferred to Nytran membranes. Filters were hybridized to ^{32}P-labeled DNA fragments derived from clones cZNF2.2, cZNF41.1, cHZF3.1, cHZF7.1 at 42°C in 50% formamide, 5×SSC, 5×Denhardt's, 0.5% SDS and washed in 0.1×SSC, 0.2% SDS at 68°C.

RESULTS

A second conserved protein module is present in several cDNA clones potentially encoding finger polypeptides

In Figure 1A part of the nucleotide sequence of a 3.923 bp long cDNA clone (cZNF2.2), homologous to the human ZNF2 (A1.5) finger gene (47), is reported. cZNF2.2 was isolated from a cDNA library from undifferentiated human teratocarcinoma NTera2-D1 cells, using as probe a DNA segment from a previously isolated, shorter cDNA clone (cZNF2.1; see Materials and Methods and below). Conceptual translation of the cZNF2.2 sequence predicted a long ORF (428 residues) which includes 9 finger units preceded by 170 amino acids. The finger motifs exhibit features of *Krüppel*-like gene products, i.e. they conform to the general pattern $CX_2CX_3FX_5LX_2HX_3H$ and are preceded by a conserved sequence of seven amino acids, the H/C link (consensus TGEKPYe; 14) The first in frame ATG within the cZNF2.2 ORF is found at nucleotides 523−525.

Visual inspection of sequences revealed that regions similar to amino acids 16−56 and 48−76 of the cZNF2.2 ORF were present in the murine Zfp1 (19) and human HPF9 (23) finger proteins, respectively. This finding suggested that a subset of these proteins may share amino acids similarities outside the finger domain. A DNA fragment spanning the region that encodes amino acids 16−76 of cZNF2.2 (probe B in Figure 1A) was used to probe the DNA of 65 cDNA clones previously isolated from the same cDNA library on the basis of cross-hybridization to an H/C link containing probe (see Materials and Methods). 15 clones hybridized to probe B; of these, 10 were analyzed further and 9 shown by Southern analysis to contain different inserts (data not shown). The nucleotide sequence of the inserts of three clones (cZNF41.1, cHZF3.1, cHZF7.1) was determined. Conceptual translation revealed that an approximately 65 amino acids long stretch, which spans the region of similarity of cZNF2.2 to Zfp1 and HPF9 (shaded residues in Figure 1), was shared amongst cZNF2.2, cZNF41.1 and cHZF7.1; in cHZF3.1 the relatedness to the others was restricted to the first 42 amino acids of this block.

A sequence data base search (see Materials and Methods) identified significant homologies also in several other previously described *Krüppel*-like finger proteins. These included the human ZNF7 (24), ZNF8 (24), Kox1 (25) and ZFP36 (48) genes and Xfin (15) isolated from *Xenopus*. Sequences that exhibit similarities to the region mentioned above are reported in Figure 2. The region shared by all sequences is within the N-terminal portion of the potential polypeptides and is separated from the finger domain by amino acid stretches of variable length. In some of these, a few similarities were found (see legend to Figure 2). The motif shared by all is referred to as finger preceding box (FPB). Amino acids similarities within the FPB encompass a stretch of 65 residues in five out of eleven sequences (cZNF2.2 through to Kox1); homology is restricted to the first 42 residues in three (cHZF3.1, Zfp1 and ZFP36). On this basis, the FPB region may be viewed as composed of two modules which include residues 1−42 (FPB-A) and residues 43−65 (FPB-B). HPF9 and ZNF8 are truncated cDNA moieties (23, 24) which do not extend over the entire FPB-A region; in *Xfin* the putative initiator methionine lies within FPB-A. Amino acid identities between each member and the derived consensus sequence range from 45% to 69% in both A and B modules. The degree of similarity among sequences increases if it is taken

into account that different amino acids with similar physico-chemical characteristics (49) occur at several positions.

It has to be noticed that in the cZNF2.2 ORF the first in frame ATG is in the FPB module (nucleotides 523–525 in Figure 1A). Analysis of a distinct cDNA indicates that an alternative splicing event places an additional coding exon carrying an in frame ATG just upstream of the FPB-A module (see below and Discussion).

Expression of FPB-containing sequences in human cell lines

In Figure 3A are shown the results of Northern blot analyses with poly(A)$^+$ RNA from undifferentiated human teratocarcinoma NTera2-D1 cells probed at high stringency with DNA fragments including the FPB regions of clones cZNF2.2, cZNF41.1, cHZF7.1 and cHZF3.1. The probes detected either a predominant (cHZF3.1; lane 4) or multiple transcripts of similar abundance (cZNF2.2, cZNF41.1, cHZF7.1; lanes 1–3). RNA sizes ranged from approximately 7.000 (top band with cHZF7.1; lane 3) to 2.600 nucleotides (bottom band with cZNF2.2; lane 1). Transcripts of similar lengths were detected with the cZNF2.2, cHZF3.1 and cHZF7.1 FPB-containing probes also in poly(A)$^+$ RNA preparations from several other cell lines (Figure 3B), indicating that these genes are expressed in various cells. Similar results were obtained with the cZNF41.1 FPB-probe (not shown). This gene was previously shown to be expressed in several cell lines (26).

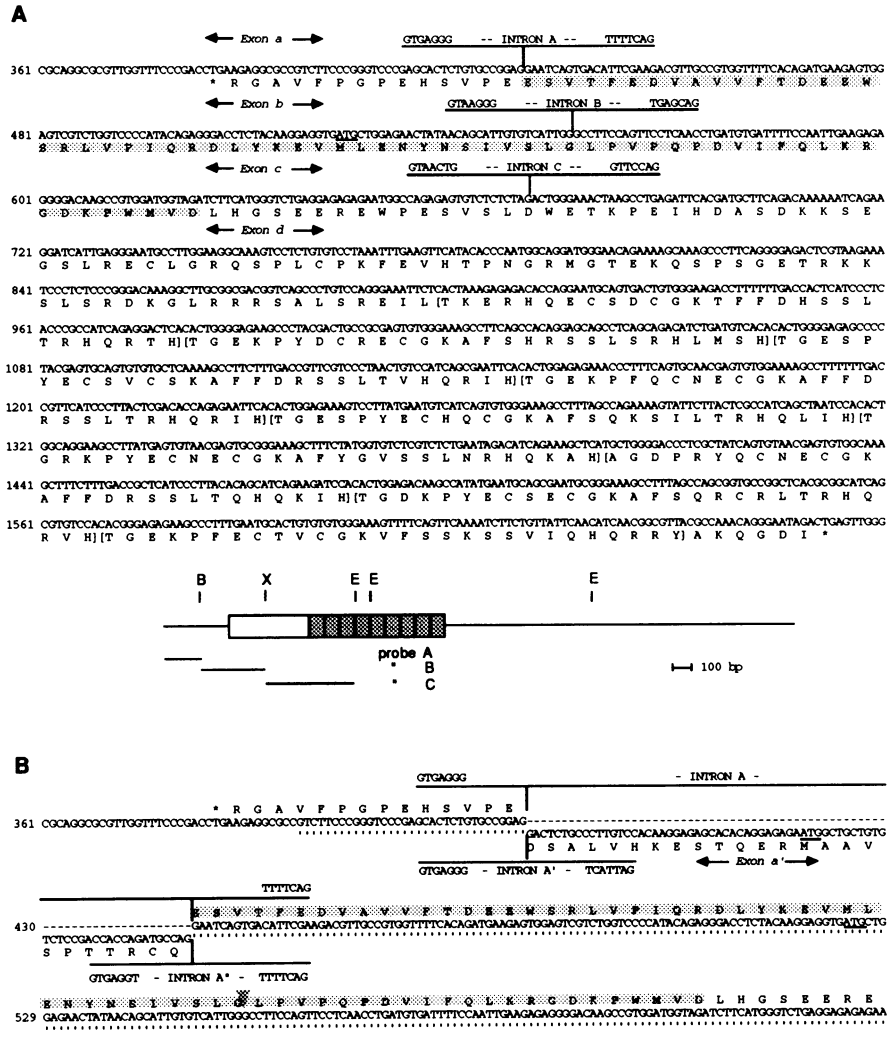

Figure 1. Nucleotide and deduced amino acid sequence of alternative cDNAs derived from the ZNF2 gene and intron-exon organization of the gene. A) Part of the nucleotide sequence of the cZNF2.2 insert and the amino acid sequence of the hypothetical zinc fingers containing ORF are shown. Numbers on the left refer to nucleotide positions (total length 3.923 bp). Asterisks denote in frame translation termination codons. The first in frame ATG codon is underlined. Brackets delimit the 9 finger motifs. The position of the three introns (A, B and C) and the sequences present at the junctions with the four exons (*a*, *b*, *c*, *d*) in the corresponding ZNF2 gene are reported. Shading highlights the amino acid stretch also found in other potential finger proteins (see text). A schematic representation of the cZNF2.2 cDNA insert is drawn at the bottom. The ORF is shown as a box. Finger repeats are delimited by bars and shaded. Restriction sites: B, BamHI; E, EcoRI; X, XbaI. The position of probes A, B and C used for library screenings and Northern analyses is also shown. Probe C corresponds to the DNA segment derived from cZNF2.1 utilized to isolate cZNF2.2 (see text and Figure 1B). B) Part of the nucleotide sequence of cZNF2.1 (bottom line; total length 824 bp) is aligned to that of cZNF2.2 (top line). Hyphens denote nucleotide identity. Conceptual translations are reported above and below each sequence, respectively. The first in frame ATG found in each is underlined. Numbers on the left refer to cZNF2.2 nucleotides as in Figure 1A; dashes mark the missing exon *a'*. The position and sequence termini of introns A, A', A'' are reported. The triangle denotes the position of intron B. The cZNF2.2 insert extends up to nucleotide 1147 of cZNF2.2.

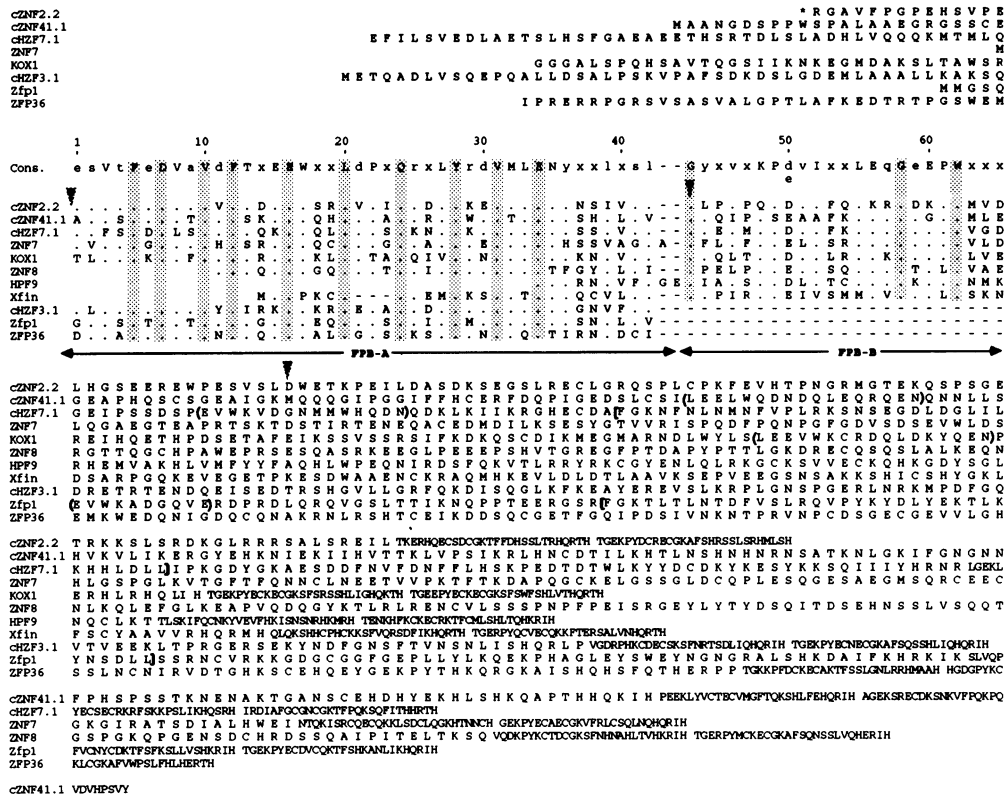

Figure 2. Alignment of deduced amino acid sequences of zinc fingers containing cDNA clones. The sequences of the N-terminal regions of potential zinc finger proteins, up to the first two finger motifs of each, are shown. Sequences are from: cZNF2.2, cZNF41.1, cHZF7.1 and cHZF3.1 (this work); ZNF7 and ZNF8 (24); Kox1 (25); HPF9 (23); Xfin(15); Zfp1(19); ZFP36 (48). A consensus sequence for the regions referred to as FPB-A and FPB-B is reported; numbering is based on the most frequent FPB lengths. X indicate positions where no residue is common to more than 50% of the sequences, uppercase residues are conserved in more than 75% of the sequences. Dots in the aligned sequences mark identity to the amino acid reported in the consensus. Shading highlights invariant residues. Triangles mark the position of introns in the ZNF2 gene. Amino acid stretches similar in cZNF41.1, cHZF7.1, Kox1, Zfp1 are in parentheses; an additional stretch similar in cHZF7.1 and Zfp1 is in brackets. The cZNF2.2 ORF is reported starting from the in frame terminator codon (marked by an asterisk, see also Figure 1 and text). Deduced amino acid sequences starting with a methionine are preceded by in frame terminators. The ORFs of cHZF7.1, Kox1 and ZFP36 may extend further upstream. Part of the reported amino acid sequence of cHZF7.1 matches that of HPF2 (amino acids 1–170; 23).

Alternative splicing and intron-exon junctions of the ZNF2 gene

In Figure 1B part of the nucleotide and predicted amino acid sequences of the insert of cZNF2.1 (824 bp), a cDNA clone isolated from a human liver library (see Materials and Methods), are aligned to those of cZNF2.2. Alignment starts at nucleotide 397 of the cZNF2.2 insert and ends at position 1147 within the finger region. The two sequences are colinear except for 72 bp. This results in distinct ORFs which differ in the absence (cZNF2.2) or presence (cZNF2.1) of 24 amino acids upstream of FPB-A (Figure 1B). The additional DNA segment in cZNF2.1 not only extends the predicted ORF but also brings in an in frame ATG codon (Figure 1B). Two bands of hybridization approximately 4.200 and 2.600 nucleotides in length are detected by Northern analysis with probes including the homologous regions of the two cDNAs (probe B, Figure 3; probe C, data not shown). Because of its length, the cZNF2.2 insert (over 3.9 kbp) likely derives from the longer transcript. The cZNF2.1 cDNA moiety (824 bp) may have originated from either RNA, as a difference in length of 72 bp may not result in distinct mobilities of RNAs in the range of 4.000 nucleotides.

Clones containing the ZNF2 gene were isolated from a cosmid library of human genomic DNA (Materials and Methods). Two overlapping cosmids (gZNF2.1 and gZNF2.2) were shown by Southern analysis to span the entire cZNF2.2 insert (data not shown). Intron-exon junctions were determined by priming sequencing reactions on the genomic DNA contained in gZNF2.1 and gZNF2.2 with a battery of oligonucleotides complementary to either the sense or anti-sense strand of cZNF2.2. The positions of the introns (A, B, C) as well as the sequences of their termini are reported in Figure 1A; their positions are also marked in Figure 2. The cZNF2.2 cDNA moiety derives from the joining of four exons (*a* through to *d*). Exon *d* contains the 3′ untranslated region (nucleotides 1672–3923; not shown) as a large portion of the potential ORF including all of the finger domain and the preceding 75 amino acid residues (Figure 1A). Exon *c* encodes the entire FPB-B module and the following 15 amino acids. The junction of intron B with exon *c* breaks the codon for the glycine residue which marks the proposed boundary between FPB-A and FPB-B (Figures 1A and 2). Exon *b* includes the FPB-A coding sequences. Exon *a* extends to the first nucleotide of the DNA insert.

The additional 72 bp found in cZNF2.1 may have been introduced through alternative splicing. This segment is indeed flanked in the ZNF2 gene by splice sites (Figure 1B); the spliced out portions of intron A are marked as A′ and A″, while the 72 bp long coding exon is referred to as *a′*.

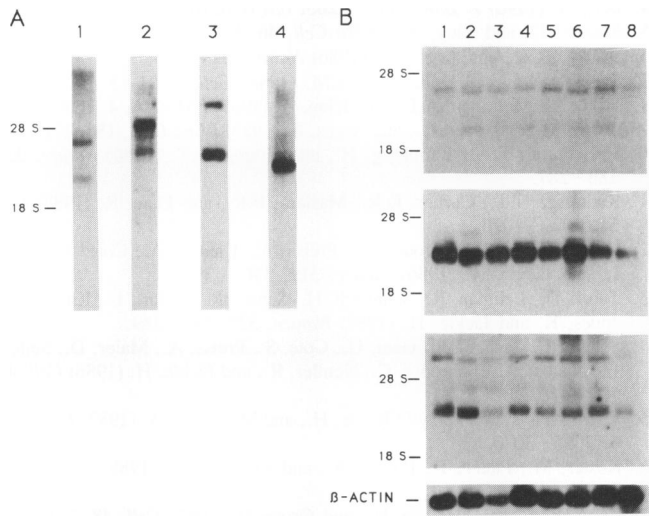

Figure 3. Expression of FPB containing finger genes in human cell lines. Five μg of poly(A)$^+$ RNA from human cell lines were processed as described under Materials and Methods for Northern analysis. (A) Poly (A)$^+$ from Ntera2-D1 cells hybridized to probes including the FPB-regions of: 1) cZNF2.2; 2) cZNF41.1; 3) cHZF7.1; 4) cHZF3.1. (B) Poly (A)$^+$ RNAs from: 1) NTera2-D1, 2) HeLa, 3) HepG2, 4) HT-1080, 5) A1251, 6) MOLT-4, 7) Jurkat, 8) HL-60. From top to bottom the filter was sequentially hybridized to the same probes from cZNF2.2, cHZF3.1, cHZF7.1 as in panel A. The filter was rehybridized to a β-actin probe. The positions of migration of 28S and 18S rRNAs are indicated.

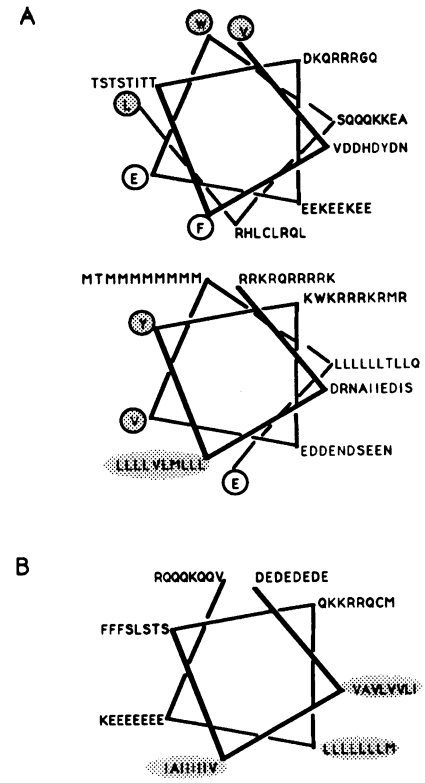

Figure 4. Helical wheel projection of three putative a-helices within FPB-A and FPB-B modules. (A) Putative helices within FPB-A. At the top helical wheel analysis of residues 10–20 (see numbering in Figure 2) at the bottom of residues 25–34. (B) A putative α-helix within FPB-B from 50 to 57. Circled amino acid correspond to invariant residues. Shading highlights position of conserved hydrophobic character. Amino acids which vary in distinct members are all reported at corresponding wheel positions; left to right typing corresponds to top to bottom presentation in Figure 2. In the two helices in A, only residues from sequences spanning entirely these regions are reported.

DISCUSSION

Finger proteins related to the *Drosophila Krüppel* gene product are encoded by large multigene families in higher eukaryotes. In humans, the number of family members was estimated to be over three hundred (23). The results reported here help to define a structurally related subset of genes which potentially encode polypeptides that share, in addition to finger motifs, a highly conserved amino acid module which encompasses approximately 65 residues. This region, referred to as FPB, invariably precedes the finger repeats from which is separated by amino acid stretches that vary in length and do not share extensive sequence homologies (Figure 2). Additional short similar motifs found within some of these linker sequences (Figure 2) may underlie distinct degrees of relatedness among family members. FPBs come in discrete different lengths as only the first 42 amino acids are present in some instances (Figure 2). This size variation indicates that the FPB is composed of two modules, FPB-A and FPB-B (amino acids 1–42 and 43–65; Figure 2). This distinction, based on sequence comparisons, finds a physical correspondance in the separation of the FPB-A and FPB-B coding sequences in two exons within the ZNF2 gene. A notable scattering of intervening sequences occurs in fact within the region spanning the FPB module in ZNF2. Three introns separate the four exons found in a cDNA representative of this gene (cZNF2.2; Figure 1A). The A and B intervening sequences flank and interrupt the first codons of the FPB-A and FPB-B modules, respectively; intron C is placed a few nucleotides downstream of the FPB-B sequences, marking their separation from the remainder of the coding region (Figures 1–2). The position of introns whithin the ZNF2 gene has interesting implications. Paralogous members of multigene families have often identically placed introns (50, 51), therefore, it may be expected that the intron-exon boundaries found in the ZNF2 gene are conserved in other FPB-family members. Coding exons often equate to distinct protein domains (52, 53); the FPB A and B modules may turn out to follow this pattern and serve distinctive functions. Also, on the basis of the intron positions within the FPB area, it can be hypothesized that alternative splicing pathways lead to the assembly of mRNAs encoding polypeptides that contain one, both or none of the FPB-A and FPB-B modules. The absence of FPB-B in cHZF3.1, Zfp1 and ZFP36 (Figure 2) may therefore be the result of splicing events which skipped the cognate exons. The utilization of distinct splicing pathways would allow the assembly of mRNAs encoding sets of polypeptides which differ in their FPB content and likely in their properties. Transcripts of different lengths are homologous to three of the FPB-containing cDNAs described here (Figure 3), a result which indicates the existence of alternative RNAs. Previously reported genes which belong to this subset (Zfp1 and ZNF7, ZNF8; Figure 2) also give rise to multiple transcripts (19, 24). Moreover, the small sizes of the FPB-A and FPB-B exons may not lead to a clear resolution of RNAs which contain both or miss either of the two.

The isolation of distinct ZNF2 homologous cDNAs indicates that alternative splicing also contributes to adjoin FPB coding sequences to variable exons. cZNF2.1 and cZNF2.2 differ in 72 bp placed, within the former, immediately upstream of FPB-A (Figure 1B). While the first methionine of the cZNF2.2 ORF

is within FPB-A, the additional 72 bp exon present in cZNF2.1 brings in an ATG codon which precedes the FPB coding sequences. Splicing events which skip or insert exon *a'* may therefore lead to the synthesis of proteins differing in length and FPB content due to alternative sites of initiation of translation. Differential splicing leading to alternative isoforms of ZFX zinc finger proteins, initiated at distinct ATGs, has been documented (54). On the other hand, it cannot be excluded that codons upstream of the first ATG are recognized as translation start sites in cZNF2.2 homologous mRNAs. Several examples indicate in fact that, though at reduced efficiency, initiation of translation may occur at triplets other than ATG (55 and references therein).

Both FPB-A and FPB-B are enriched in negatively charged amino acids (approximately 20% of the residues are aspartic or glutamic acid) and in hydrophobic leucines and valines (>20%). Invariant residues are found at thirteen positions within the FPBs analyzed (Figure 2). Data base searches did not provide indications of significant similarity to protein motifs of known structure or function. It is of interest to mention that the results of different secondary structure prediction methods on all aligned sequences (Materials and Methods) consistently suggest that three regions centered around positions 16 and 30 (FPB-A) and position 54 (FPB-B) have a high probability of α-helical folding. Helical wheel projections (Figure 4) suggest that putative α-helices within these regions would show an amphipathic character; in the two FPB-A putative helices the invariant amino acids would all be positioned on one face of the helix.

Proteins involved in the control of gene expression often include conserved modules which confer specific functional properties. Common structural motifs as finger repeats and homeoboxes contribute to provide the products of two major gene families with the ability to interact specifically with nucleic acids. The presence of additional conserved modules as the paired box (56) and the POU domain (57) defines structurally related subsets of homeoproteins with distinctive properties (58). Although it is difficult to speculate at this stage on the role of FPB modules, their high level of conservation strongly suggests that they may have functional and/or structural relevance. In the second hypothesis the presence of putative amphipathic helical motifs points at the possibility of inter or intra-molecular interactions.

Recently, the characterization of another set of human cDNA clones encoding polypeptides which contain both zinc finger repeats and a region similar to the FPB has been reported (59).

ACKNOWLEDGMENTS

We thank J.Skowronski and M.F.Singer for the gift of NTera2-D1 cDNA libraries, H.Jäckle and P.Gruss for the gift of finger probes, Chris Sander for making the COLLATE package available as well as his own secondary structure prediction program SEGMENT83 and R. Cortese for support and encouragement in the initial phases of this work. We are indebted to P.P.Di Nocera, E.Boncinelli and R.Driscoll for critical reading of the manuscript. This work was supported in part by grants from P.F.Ingegneria Genetica, C.N.R and from IV Progetto AIDS, Ministero della Sanità.

REFERENCES

1. Diakun, G.P., Fairall, L. and Klug, A. (1986) *Nature*, **324**, 698.
2. Frankel, A.D., Berg, J.M., and Pabo, C.O. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 4841–4845.
3. Parraga, G., Horvath, S.J., Eisen, A., Taylor, W.E., Hood, L., Young, E.T., and Klevit, R.E. (1988) *Science*, **241**, 1489–1492.
4. Berg, J. (1990) *J. Biol. Chem.*, **265**, 6513–6516.
5. Rhodes, D. and Klug, A. (1986) *Cell*, **46**, 123–132.
6. Brown, R.S., and Argos, P. (1986) *Nature*, **324**, 215.
7. Evans, R.M., and Hollenberg, S.M. (1988) *Cell*, **52**, 1–3.
8. Miller, J., McLahan, A.D., and Klug, A. (1985) *EMBO J.*, **4**, 1609–1614.
9. Brown, R.S., Sander, C., and Argos, P. (1985) *FEBS Lett.*, **186**, 271–274.
10. Hartshorne, T.A., Blumberg, H., and Young, E.T. (1986) *Nature*, **320**, 283–287.
11. Kadonaga, J.T., Carner, K.R., Masiarz, F.R., and Tijan, R. (1987) *Cell*, **51**, 1079–1090.
12. Rosenberg, U.B., Schroder, C., Preiss, A., Kienlin, A., Cote, S., Riede, I., and Jäckle, H. (1986) *Nature*, **319**, 336–339.
13. Tautz, D., Lehman, R., Schnurch, H., Schuh, R., Seifert, E., Kienlin, A., Jones, K. and Jäckle, H. (1987) *Nature*, **327**, 383–389.
14. Schuh, R., Aicher, W., Gaul, U., Cote, S., Preiss, A., Maier, D., Seifert, E., Nauber, U., Schroder, C., Kemler, R., and Jäckle, H. (1986) *Cell*, **47**, 1025–1032.
15. Ruiz i Altaba, A., Perry-O'Keefe, H., and Melton, D.A. (1987) *EMBO J.*, **6**, 3065–3070.
16. Koster, M., Pieler, T., Poting, A., and Knochel, W. (1988) *EMBO J.*,**7**, 1735–1741.
17. Chowdhury, K., Deutsch, U., and Gruss, P. (1987) *Cell*, **48**, 771–778.
18. Chavrier, P., Lemaire, P., Relevant, O., Bravo, R., and Charnay, P. (1988) *Mol. Cell. Biol.*, **8**, 1319–1326.
19. Chowdhury, K., Dietrich, S., Balling, R., Guenet, J.L. and Gruss, P. (1989) *Nucleic Acids Res.*, **17**, 10427–10438.
20. Cunliffe, V., Koopman, P., Mc Laren, A. and Trowsdale, J. (1990) *EMBO J.*, **9**, 197–205.
21. Ruppert, J.M., Kinzler, K.W., Wong, A.J., Bigner, S.H., Kao, F.T., Law, M.L., Seunaez, H.N., O'Brien, S.J., and Vogelstein, B. (1988) *Mol.Cell. Biol.*, **8**, 3104–3113.
22. Pannuti, A., Lanfrancone, L., Pascucci, A., La Mantia, G., Pellicci, P.-G. and Lania, L. (1988) *Nucleic Acids Res.*,**16**, 4227–4237.
23. Bellefroid, E.J., Lecocq, P.J., Benhida, A., Poncelet, D.A., Belayew, A., and Martial, J. A. (1989) *DNA*, **8**, 377–387.
24. Lania, L. Donti, E., Pannuti, A., Pascucci, A. Pengue, G., Feliciello, I., La Mantia, G., Lanfrancone, L., and Pelicci, P-G. (1990) *Genomics*, **6**, 333–340.
25. Thiesen, H.J. (1990) *New Biol.*, **2**, 363–374.
26. Franzé, A. M., Archidiacono, N., Rocchi, M., Marino, M. and Grimaldi, G. (1991) *Genomics*, **7**, 728–736.
27. Chavrier, P., Zerial, M., Lemaire, P., Almendral, J., Bravo, R. and Charnay, P. (1988) *EMBO J.*, **7**, 29–35.
28. Lemaire, P., Relevant, O., Bravo, R. and Charnay, P. (1988) *Proc. Natl. Acad., Sci. USA*, **85**, 4691–4695.
29. Sukhatme, V.P., Cao, X., Chang, L.C., Tsai-Morris, C.H., Stamenkovic, D., Ferreira, P., Cohen, D.R., Edwards, S.A., Shows, T.B., Curran, T., LeBeau, M.M., and Adamson, E.D. (1988) *Cell*, **53**, 37–43.
30. Joseph, L.J., Le Beau, M.M., Jamieson, G.A., Acharya, S., Shows, T.B., Rowley, J.D., and Sukhatme, V.P. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 7164–7168.
31. Kinzler, K. W., Ruppert, J. M., Bigner, S.H. and Vogelstein, B. (1988) *Nature*, **332**, 371–374.
32. Morishita, K., Parker, D.S., Mucenski, M. L., Jenkins, N.A., Copeland, N. G. and Ihle, J. N. (1988) *Cell*, **54**, 831–840.
33. Call, K.M., Glaser, T., Caryn, Y.I., Buckler, A.J., Pellettier, J., Haber, D.A., Rose, E.A., Kral, A., Yeger, H., Lewis, W.H., Jones, C., and Housman, D.E. (1990) *Cell*, **60**, 509–520.
34. Knochel, W., Poting, A., Koster, M., El Baradi, T., Nietfeld, W., Bouwmeester, T., and Pieler, T. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 6097–6100.
35. Mc Alpine, P.J., Shows, T.B., Boucheix, C., Stranc, L.C., Berent, T.G., Pakstis, A.J. and Douté, R.C. (1989) *Cytogenetics and Cell Genetics*, **51**, 13–66.
36. Skowronski, J., Fanning, T.G. and Singer, M.F. (1985) *Mol. Cell. Biol.*, **8**, 1385–1397.
37. Grimaldi, G., Manfioletti, G. and Schneider, C. (1987) *Nucleic Acids Res.*, **15**, 9608.
38. Huynh, T.V., Young, R.A. and Davis, R.W. (1988) In Glover, D. M. (ed.) DNA Cloning-A practical approach. IRL Press, Oxford, Vol. I, pp. 49–78.
39. Poustka, A., Rackwitz, H. R., Frischauf, A.M., Hohn, B. and Lehrach, H. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4129–4133.
40. Maniatis,T., Fritsch, E.F. , and Sambrook, J. (1982) Molecular Cloning: A laboratory manual. Cold Spring Harbor University Press, Cold Spring Harbor.
41. Manfioletti, G. and Schneider, G. (1988) *Nucleic Acids Res.*, **16**, 2873–2884.

42. Devereux, J., Haeberli, P. and Smithies, O. (1984) *Nucleic Acids Res.*, **12**, 387–395.
43. Argos, P. (1987) *J. Mol. Biol.*, **193**, 385–396.
44. Garnier, J., Osguthorpe, and Robson, B. (1978) *J. Mol. Biol.*, **120**, 97–120.
45. Gibrat, J., Garnier, J. and Robson, B., (1987) *J. Mol. Biol.*, **198**, 425–443.
46. Ptitsyn, O.B. and Finkelstein, A.V. (1983) *Biopolymers*, **22**, 15–25.
47. Archidiacono, N., Grimaldi, G., Rocchi, M. and Romeo, G. (1989) *Cytogenet. Cell Genet.*, **51**, 952.
48. Cunliffe, V. and Trowsdale, J. (unpublished). S08686.
49. Mc Keon, F.D., Kirschner, M.W. and Caput, D. (1986) *Nature*, **319**, 463–468.
50. Gilbert, W. (1985) *Science*, **228**, 823–824.
51. Blake, C. C. F., (1985) *Int. Rev Cytol.*, **93**, 149–185.
52. Collins, F., S. and Weissmann, S. M. (1984) *Progr. Nucl. Acid Res. Mol. Biol.*, 31, 317–438.
53. Wilde, C.D., Chow, L.T., Wefald, F.C. and Cowan, N.J. (1982) *Proc. Natl. Acad. Sci.* USA, **79**, 96–100.
54. Schneider-Gèdicke, A., Beer-Romero, P., Brown, L.G., Mardon, G., Luoh, S.-W., and Page, D.C. (1989) Nature, 342, 708–710.
55. Saris, C., J.M., Domen, J. and Berns, A. (1991) *EMBO J.*, **10**, 655–664.
56. Bopp, D., Burri, M., Baumgarter, S., Frigerio, G., and Noll, M. (1986) *Cell*, **47**, 1033–1040.
57. Herr W., Sturm, R.A., Clerc, R.G., Corcoran, L.M., Baltimore, D., Sharp, P.A., Ingraham, H.A., Rosenfeld, M.G., Finney, M., Ruvkun, G., and Horvitz, H.R. (1988) *Genes Dev.*, **2**, 1513–1516.
58. Ruvkun, G. and Finney, M. (1991) *Cell*, **64**, 475–478.
59. Bellefroid, E.J., Poncelet, D. A., Lecocq, P.J., Relevant, O. and Martial, J.A. (1991) *Proc. Natl. Acad. Sci.* USA, **88**, 3608–3612.