Cloning and Sequencing of Rat Plectin Indicates a 466-kD Polypeptide Chain with a Three-Domain Structure Based on a Central Alpha-Helical Coiled Coil

Gerhard Wiche,* Bruno Becker,* Karin Luber,* Georg Weitzer,* Maria J. Castañon,‡ Rudolf Hauptmann,‡ Christian Stratowa,‡ and Murray Stewart§

*Institute of Biochemistry, University of Vienna, 1090 Vienna, Austria; ‡Ernst Boehringer Institut, 1121 Vienna, Austria; and §MRC Laboratory of Molecular Biology, Cambridge CB2 2QH, England

Abstract. We have determined the complete cDNA sequence of rat plectin from a number of well-characterized overlapping lambda gt11 clones. The 4,140-residue predicted amino acid sequence (466,481 D) is consistent with a three-domain structural model in which a long central rod domain, having mainly an alpha-helical coiled coil conformation, is flanked by globular NH_2- and COOH-terminal domains. The plectin sequence has a number of repeating motifs. The rod domain has five subregions ~200-residues long in which there is a strong repeat in the charged amino acids at 10.4 residues that may be involved in association between plectin molecules. The globular COOH-terminal domain has a prominent six-fold tandem repeat, with each repeat having a strongly conserved central region based on nine tandem repeats of a 19-residue motif. The plectin sequence has several marked similarities to that of desmoplakin (Green, K. J., D. A. D. Parry, P. M. Steinert, M. L. A. Virata, R. M. Wagner, B. D. Angst, and L.A. Nilles. 1990. *J. Biol. Chem.* 265:2,603–2,612), which has a shorter coiled-coil rod domain with a similar 10.4 residue charge periodicity and a COOH-terminal globular domain with three tandem repeats homologous to the six found in plectin. The plectin sequence also has homologies to that of the bullous pemphigoid antigen. Northern blot analysis indicated that there is a significant degree of conservation of plectin genes between rat, human, and chicken and that, as shown previously at the protein level, plectin has a wide tissue distribution. There appeared to be a single rat plectin gene that gave rise to a 15-kb message. Expression of polypeptides encoded by defined fragments of plectin cDNA in *E. coli* has also been used to localize the epitopes of a range of monoclonal and serum antibodies. This enabled us to tentatively map a sequence involved in plectin–vimentin and plectin–lamin B interactions to a restricted region of the rod domain.

PLECTIN is an abundant high molecular weight cytomatrix protein found in a wide variety of different cell types. Based mainly on biochemical and immunolocalization studies it has been proposed that plectin plays a role in the cross-linking of intermediate filaments, the interlinking of intermediate filaments with microtubules and microfilaments, and the anchoring of intermediate filaments to the plasma membrane and the nuclear membrane (for reviews see Wiche, 1989; Foisner and Wiche, 1991.). Furthermore, plectin's ability to form networks through self-association may convey structural stability to cytoplasmic areas which are devoid of cytoskeletal filaments. Plectin molecules exist in solution predominantly as dumbbell-shaped ~200-nm-long structures that have a high tendency to aggregate via their globular end domains yielding networklike protein assemblies (Foisner and Wiche, 1987). The rod section has been suggested to play a role in plectin–vimentin and plectin–lamin B interactions (Foisner et al., 1988; 1991*a*,*b*). The location of binding sites of other known interaction partners of plectin, such as the microtubule-associated proteins MAP 1 and 2, and the membrane skeleton proteins, α-spectrin, and fodrin (Herrmann and Wiche, 1987) is largely unknown.

To elucidate the primary structure of plectin and to begin to study questions regarding the domain organization, expression, and proposed functions of the protein in greater detail, we have isolated cDNAs encoding plectin from rat and man. We report here the complete nucleotide sequence of the coding region of rat plectin and make predictions regarding plectin's molecular structure based on computer-aided analyses. Furthermore, we discuss implications of structural homologies between rat plectin and human desmoplakin that were detected by this study.

1. *Abbreviation used in this paper*: BP, bullous pemphigoid.

Materials and Methods

Antibodies

Rabbit antibodies to plectin (Wiche and Baker, 1982) were affinity purified by adsorption to a sample of purified rat glioma C_6 cell plectin immobilized on Affi-gel 15 (Bio-Rad Laboratories, Richmond, CA) and by elution with 0.2 M glycine-HCl, pH 2.5, and 0.15 M sodium chloride. mAbs to plectin were prepared as described elsewhere (Foisner et al., 1991*a*).

Construction of cDNA Library

To construct a cDNA library in lambda gt11, total RNA was isolated from cultured rat glioma C_6 cells by extraction with guanidinium thiocyanate at low temperature (Han et al., 1987) and poly(A)$^+$ RNA purified by chromatography on oligo-dT cellulose (type 3; Collaborative Research Inc., Lexington, MA) (Aviv and Leder, 1972). Double stranded, oligo-dT-primed cDNA was synthesized according to Gubler and Hoffman (1983) using a commercial cDNA synthesis system (Amersham International, Amersham, UK) except that the incubation period for the synthesis of the first strand was 2 h, and the reverse transcriptase and the ribonuclease inhibitor supplied by the manufacturers were replaced by Super RT (AMV reverse transcriptase; P.H. Stehelin & Cie AG, Basel, Switzerland) and RNAsin (Promega Biotec, Madison, WI), respectively. After EcoRI methylase treatment, EcoRI linkers were added and the cDNA subsequently digested with EcoRI to restore cohesive ends. Excess linkers were removed by gel filtration on a Fractogel TSK HW 55 (F) (E. Merck, Darmstadt, Germany) column (40 cm × 0.5 cm). The cDNA (150 ng) was concentrated by ethanol precipitation and ligated to 1.5 μg of EcoRI-digested, dephosphorylated lambda gt11. The recombinant phages were packaged in vitro using Gigapack Plus (Stratagene Inc., La Jolla, CA) and then plated onto *Escherichia coli* strain Y1088.

Isolation of cDNA Clones

The lambda gt11 library was screened with antiserum to plectin (Wiche and Baker, 1982) using the Proto-Blot detection system (Promega Biotec, Madison, WI). To reduce the background signal, the antiserum was preincubated with 1 mg/ml of *E. coli* proteins for 30 min. Before screening, plaque lifts on isopropyl-β-D-thiogalacto-pyranoside (IPTG)l-saturated nitrocellulose filters were blocked overnight in 20% FCS, 0.05% Tween 20 in TBS, pH 8. Blocked filters were submerged for 1 h in preadsorbed, primary antiserum diluted 1:500 in TBS plus 0.05% Tween 20. Bound antibodies were detected with alkaline phosphatase–conjugated antibodies to rabbit immunoglobulins and subsequent color development. Antibody reactive clones were identified and plaque purified. Their cDNA inserts were isolated and subcloned into plasmid pUC 8 (Vieira and Messing, 1982). Additional screening was performed using radiolabeled (see below) cDNA probes derived from various clones (see text). In these cases the nitrocellulose filters were prehybridized for 2 h at 42°C in 50% deionized formamide, 5× SSC, 10× Denhard's solution (0.2% Ficoll 400, 0.2% polyvinylpyrollidone, 0.2% BSA), 0.1% SDS, 30 mM sodium phosphate, pH 7, 0.1% sodium pyrophosphate, and 100 μg/ml denatured, sonicated salmon sperm DNA. For hybridization, cDNA probes were radiolabeled with [γ-^{32}P]dCTP to a sp act of ~2×10^8 cpm/mg (Feinberg and Vogelstein, 1983) and then diluted with freshly prepared prehybridization solution to yield a sp act of $1 \times 10^6 - 1 \times 10^7$ cpm/ml. Hybridization was carried out overnight at 42°C. Unbound probe was removed by washing the filters at a final stringency of 0.2× SSC and 0.1% SDS at 50°C for 30 min.

DNA Sequencing and Computer Analysis

Plectin cDNA from clone C1 was subcloned into linearized Bluescript-KS(+) vector (Stratagene Inc.) to generate nested deletions by exonuclease III–mung bean nuclease treatment (Henikoff, 1984). cDNAs from all other clones to be sequenced, except for C7, were sonicated, end polished with the Klenow fragment of *E. coli* polymerase I, and subcloned into the SmaI site of phage M13 mp18 (Messing, 1983). In addition, the full-length inserts and any isolated restriction fragments of these clones were inserted into the EcoRI, or other appropriate cloning sites, respectively, of the same phage. Clone C7 was sequenced after subcloning a series of restriction fragments in phage M13 mp18. The internal EcoRI site of clones C7 and C8 was verified by directly sequencing the amplified insert of clone C8. Preparation of the lambda DNA template and PCR amplification between two primers flanking the cloning site of lambda gt11 were done according to Rasmussen et al. (1989). DNA sequencing was performed by the dideoxy chain-termination method (Sanger et al., 1977). Sequences were generated from both strands; ambiguous regions were further resolved using oligonucleotide primers corresponding to nearby sequences. Sequencing artifacts generated as the result of G/C compressions were avoided by using the nucleotide analogue deoxyinosine-triphosphate (dITP). Internal primers were prepared using an automated DNA synthesizer (Applied Biosystems, model 381A). Computer-aided assembly of the DNA sequences was performed using the program of Staden (1982). Nucleotide and derived amino acid sequences were analyzed using software packages from the University of Wisconsin Genetic Computer Group (UWGCG, Madison, WI), DNA Star, Inc. (Madison, WI) and MacMolly (Soft Gene, Berlin, Germany). Dot matrix comparisons were performed with the programs COMPARE and DOTPLOT (MacMolly and UWGCG; Maizel and Lenk, 1981). Alignments of homologous regions and derived computations (percent of identical and similar amino acids; Dayhoff, 1978) were done using the programs GAP (UWGCG) and ALIGN (DNA Star Inc.; Lipman and Pearson, 1985). Basic protein parameters were obtained with the program PROTEIN (DNA Star Inc.). Secondary structure predictions were analyzed according to the methods of Chou and Fassman (1978) and Robson and Suzuki (1976) via the programs PEPTIDESTRUCTURE and PLOTSTRUCTURE (UWGCG; Emini et al., 1985). The antigenetic index was calculated according to the algorithm of Jameson and Wolf (1988). Fourier transforms were calculated as described by McLachlan and Stewart (1976). The chemical sequence of the protein was translated into a mathematical sequence by scoring each residue type appropriately (thus, for example, to investigate the acidic residues, they would be scored as one and all other residues as zero). Fourier transforms were scaled so that the intensity of peaks was proportional to the negative logarithm of the probability that such a value could be produced by chance (see McLachlan and Stewart, 1976). Fourier transforms of secondary structure potentials were computed as described by Noegel et al. (1989). Database searches were done with the program FASTA (Pearson and Lipman, 1988) using the EMBL and SWISSPROT databanks.

Southern and Northern Blot Analyses

Genomic DNA (3 μg) from rat glioma C_6 cells was digested with restriction endonucleases and the fragments subjected to electrophoresis and transfer to Hybond nylon membranes (Amersham International) by capillary blotting (Reed and Mann, 1985). Prehybridization of membranes, hybridization of radiolabeled cDNA probes to the fragments, and removal of unbound probes were performed as described above.

RNA from rat tissues, human placenta, human fibrosarcoma cells (Hs 913T), and chicken embryo fibroblasts was isolated similar to rat glioma C_6 cell RNA (see above). Samples of total RNA (10 μg) or poly(A)$^+$ RNA (2 μg) from different sources were denatured by heating to 65°C for 5 min in a solution containing 50% formamide (deionized), 2.3 M formaldehyde (deionized), and subsequent cooling on ice. RNA species were separated on vertical 1 or 1.5% agarose gels in 2.2 M formaldehyde (deionized), 5 mM sodium acetate, 1 mM EDTA, and 10 mM sodium phosphate, pH 7. To facilitate the transfer of large RNA species, the RNA was partially hydrolyzed by soaking the gel in alkali (Maniatis et al., 1982). After neutralizing the gel with 0.1 M Tris-HCl, pH 7.5, the RNA was transferred to nitrocellulose by capillary blotting in 20 × SSC. Prehybridization, hybridization, and removal of unbound probe were carried out as described above. Autoradiography was usually done for 8 d using an intensifying screen. To monitor the relative amounts of RNAs blotted, blots were stained with methylene blue (Zinn et al., 1983). Autoradiograms and photographic replicas of the methylene blue–stained filters were scanned using a transmission scanning device (Elscript 400, Hirschmann). ^{32}P-end-labeled HindIII restriction fragments of lambda DNA or a mixture thereof with radiolabeled HaeIII fragments of ØX174 RF DNA were used as size markers in all blot analyses.

Expression of cDNA-encoded Proteins and Immunoblotting

Two expression systems were used: (*a*) Lysogens were generated by infecting *E. coli* strain Y-1089 with phages from immunopositive plaques (Huynh et al., 1985) and cDNA-encoded proteins, fused to β-galactosidase, were produced upon induction with IPTG. (*b*) cDNA-encoded proteins fused to no more than nine vector-derived amino acids were expressed after subcloning of cDNAs into the plasmids pIMS1, pIMS5, and pIMS6 (Simon et al., 1988) and transformation of *E. coli* strain XL-1 blue (Stratagene Inc.). Ampicillin-resistant bacterial colonies expressing plectin epitopes were identified by immunoscreening (serum dilution, 1:1000) following the

method of Simon et al. (1988), except that the bacterial colonies were lysed by freezing in liquid nitrogen.

To analyze expressed proteins, bacterial extracts (Huynh et al., 1985; Simon et al., 1988) were subjected to SDS 7.5% PAGE, and the separated proteins transferred electrophoretically to nitrocellulose membranes. Blots were incubated with antibodies to plectin, at concentrations indicated, followed by secondary antibodies and the color reagents supplied with the Protoblot immunoscreening system (see above).

Results

Isolation and Verification of cDNA Clones Encoding Plectin from Rat and Man

An oligo dT-primed lambda gt11 cDNA library was constructed using poly(A)$^+$RNA prepared from rat glioma C_6 cells. A library of 2.8×10^6 independent recombinants was obtained, 95% of which contained an insert. The library was screened with an antiserum to plectin that had been preabsorbed with *E. coli* proteins to reduce background. Screening of 6×10^5 plaques yielded eight immunoreactive clones, only one of which encoded a fusion protein that retained its immunoreactivity after probing with affinity-purified antibodies to plectin. This clone, termed C1, contained an insert of 4.3 kb (Fig. 1). In a similar screening of a lambda gt11 cDNA library prepared from human placenta (Maurer-Fogy et al., 1988), nine immunopositive clones were identified, two of which (clones HP1 and HP2) retained their immunoreactivity towards affinity-purified antibodies to plectin. Clones HP1 and HP2, which had cDNA inserts of 1.5 and 0.7 kb, respectively, hybridized to each other as well as to clones C1 and C2 (Fig. 1).

Rat cDNA clones extending the C1 sequence were isolated by rescreening (5×10^5 plaques) the glioma C_6 library with a radiolabeled cDNA probe (whole insert of C1). Selection of clones C2, C3, and C4 (Fig. 1) from a large number of hybridizing clones was based on their length and degree of overlap with clone C1, as established by restriction mapping

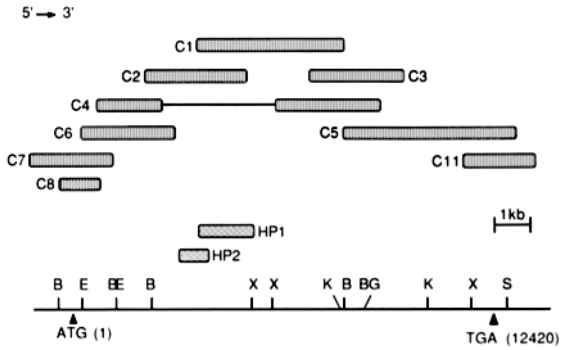


Figure 1. Alignment of overlapping cDNA clones. Clones labeled *C* were isolated from a rat glioma C_6 cell cDNA library, those labeled with *HP* from a similar human placenta library (see text). Note that clone C4 (~5 kb) was continuous, but lacked ~3.3 kb of internal sequence found in clones C1, C2, and C3 combined; it is not known whether clone C4 represented a cloning artifact or a differential splicing product. A large number of additional clones that were isolated and characterized, but were irrelevant for this presentation are not shown. The location of restriction sites mentioned in the text is indicated: *B*, BamHI; *BE*, BstEII; *BG*, BglII; *K*, KpnI; *S*, SspI; *X*, XhoI.

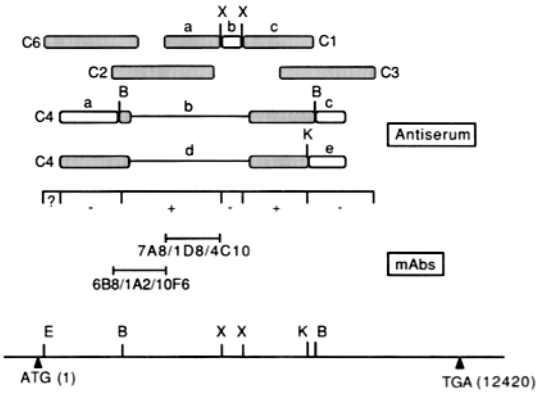


Figure 2. Expression of serum and mAb epitopes. Bars in upper part correspond to cDNA clones or restriction fragments derived from them, whose expression led to immunopositive (▨) or immunonegative (□) polypeptides. Areas encoding or not encoding epitopes are marked with + and −, respectively. Brackets in lower part mark regions encoding mAb epitopes. Note, that mAb listing does not reflect their relative position within regions.

and Southern blot hybridization using selected fragments of clone C1 as probes. In further rounds of screening, clone C5 was isolated using a 400-bp BglII–EcoRI fragment derived from the 3′ end of clone C4; clones C6, C7, and C8 were obtained by screening with a mixture of two fragments (100- and 350-bp long) derived from the 5′ ends of clones 2 (EcoRI–BamHI) and C4 (EcoRI–BstEII). Screening with an EcoRI–SspI fragment (350 bp) derived from the 3′ end of clone C5 led to the isolation of clone C11.

To confirm that indeed plectin had been cloned, we performed a series of experiments involving multiple plectin-specific immunoreagents. First, an antiserum to plectin was used to assess the immunoreactivity of cDNA-encoded proteins by colony immunoblotting after subcloning cDNAs, which had been isolated in three independent rounds of screening, in the IPTG-inducible expression vector system pIMS (Simon et al., 1988). The cDNA inserts of the five clones tested (Fig. 2) together spanned over ~9 kb, corresponding to 3/4 of the open reading frame established by sequencing (see below). All the proteins encoded by these clones were found to be immunoreactive. Positive reactions were also observed after expression of proteins encoded by two restriction fragments of clone C1 (a and c) and of clone C4 (b and d). On the other hand, the polypeptide encoded by a central XhoI fragment of clone C1 and three proteins encoded by peripheral restriction fragments of clone C4 (a, c, and e) were not immunoreactive.

In addition, a series of mAbs recognizing distinct epitopes of the plectin molecule (Foisner et al., 1991*a*; and unpublished data) was used to probe cell lysates of two independent clones (C1 and C2) by Western blot analysis. A number of these mAbs (7A8, 1D8, and 4C10) were found to be immunoreactive with both clone C1- and clone C2-encoded proteins (Fig. 3; and data not shown), whereas others (mAbs 1A2, 6B8, and 10F6) only showed positive reactions with clone C2-encoded proteins (Fig. 3). As expected, when the proteins encoded by each of three subfragments of clone C1 (Fig. 2) were examined individually using mAbs 1D8 and 7A8, only fragment *a* protein, corresponding to the region

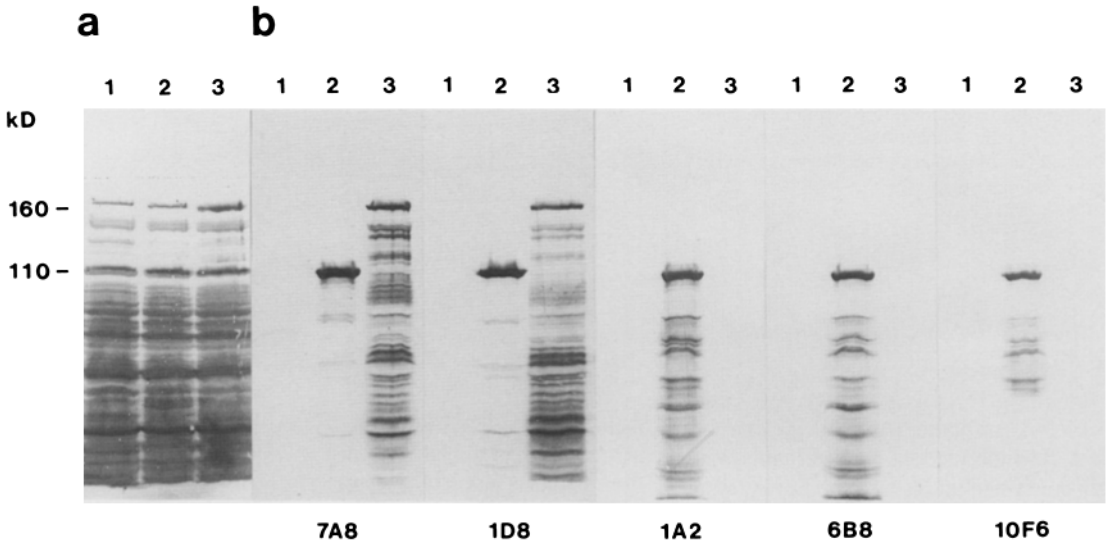


Figure 3. Western blotting of proteins expressed in pIMS vector using a series of mAbs to plectin. Lanes *1*, vector without insert; lanes *2*, clone C2 insert; lanes *3*, clone C1 insert. *a*, India ink-stained blot. *b*, immunoblot using mAbs indicated. For antibody overlays hybridoma supernatants were used at dilutions of 1:1; anti-mouse Ig's conjugated to alkaline phosphatase were used as secondary antibodies. Note that mAbs 1D8 and 7A8 were immunoreactive with proteins encoded by both clone C1 and clone C2 cDNAs, whereas mAbs 1A2, 6B8, and 10F6 detected only the proteins encoded by clone C2 cDNA.

overlapping with clone C2, was found to be immunoreactive (data not shown). Based on this type of analysis, the epitopes of these mAbs could be grouped and roughly assigned to two different areas (Fig. 2) in accordance with epitope maps obtained by EM and partial proteolytic digestion of the protein (Foisner et al., 1991*a*). These data clearly demonstrated that the isolated cDNAs encoded multiple epitopes of the plectin molecule and that the localization of these epitopes as revealed by expression of recombinant proteins was fully consistent with data obtained on the protein level.

Furthermore, when polypeptides encoded by the human cDNA clones HP1 and HP2 were expressed in *E. coli* and cell lysates analyzed by SDS-PAGE, the β-galactosidase fusion proteins generated (170 and 140 kD, respectively), were immunoreactive with affinity-purified antibodies to plectin (Fig. 4). Moreover, in a retroblot experiment, plectin antibodies, bound to and eluted from nitrocellulose-immobilized 170-kD fusion protein of clone HP1, maintained their ability to bind to plectin (data not shown).

Size, Conservation, and Tissue Distribution of Plectin mRNA

When Northern blots of total RNA from rat heart were probed with the cDNA insert of clone C1, a single mRNA species of ~15 kb was identified (Fig. 5 *a*). However, because of the poor resolution of such large RNAs in agarose gels, the existence of more than one plectin transcript of ~15 kb as a result of alternative RNA processing would probably not have been detected. To investigate the conservation of plectin genes among different species, equal amounts of poly(A)$^+$ RNAs prepared from rat glioma C_6 cells or human placenta were probed with plectin cDNAs derived from both the human (clone HP1) and the rat (clone C1) lambda gt11 libraries. Both plectin probes hybridized to homologous as well as heterologous mRNAs of ~15 kb, clearly revealing a structural conservation of plectin genes in man and rat. However, the hybridization signals were significantly stronger for the homologous, compared to the heterologous systems (Fig. 5 *b*; and data not shown). When preparations of total RNA from human fibrosarcoma (Hs 193T) cells and from chicken embryo fibroblast cells were probed with the

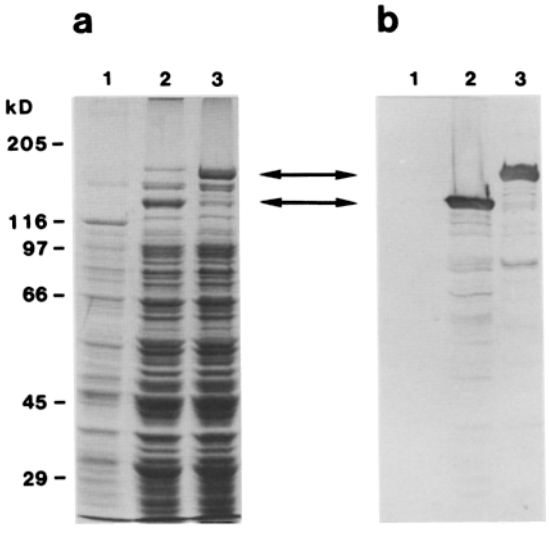


Figure 4. Expression of immunoreactive fusion proteins encoded by human placenta cDNAs. Lysates were prepared from *E. coli* strain Y1089 infected with clones HP1 and HP2 and proteins separated on a SDS 7.5% polyacrylamide gel. *a*, Coomassie blue-stained gel; *b*, corresponding immunoblot using affinity-purified antibodies to plectin (diluted 1:1,000, compared to serum). Arrows denote intact fusion proteins. Lane *1*, wild type lambda gt11; lanes *2* and *3*, lambda gt11 with clone HP2 and clone HP1 inserts, respectively. Molecular mass standards are indicated.

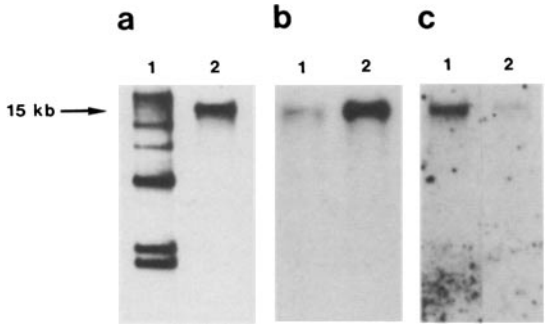


Figure 5. Northern blot analysis of RNAs isolated from three different species. *a*, Lane *1*, size markers; lane *2*, rat heart total RNA (10 μg) hybridized with a rat plectin probe (clone C1). *b*, Equal amounts (2 μg) of poly(A)$^+$ RNAs from rat glioma C_6 cells (lane *1*) and human placenta (lane *2*) hybridized with a human plectin probe (clone HP1). *c*, Equal amounts (10 μg) from human fibrosarcoma cells Hs 913T (lane *1*) and chicken embryo fibroblast cells (lane *2*) hybridized with a rat plectin probe (clone C1). Size markers (*a*, lane *1*): HindIII-cut lambda DNA (23.1, 9.4, 6.6, 4.4, 2.3, and 2.0 kb).

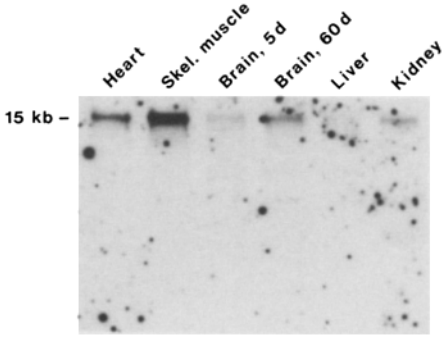


Figure 6. Tissue distribution of plectin mRNA. Poly(A)$^+$ RNAs (2 μg) from various dissected tissues of adult (60-d-old) rat and brain of newborn (5-d-old) rat were subjected to Northern blot analysis and probed with rat plectin cDNA (clone C1). The relative levels of plectin mRNA in these tissues, normalized to the total amount of RNA blotted, were estimated as follows: *skeletal muscle*, 100; *heart*, 60; *60-d-old brain*, 50; *kidney*, 30; *5-d-old brain*, 10; and *liver*, 2. Position of 15-kb DNA size marker is indicated.

rat cDNA (clone C1), chicken RNA only cross-hybridized weakly compared to human RNA (Fig. 5 *c*). These differences indicated that the primary structure of plectin transcripts, although conserved, varied among species, in particular between chicken and mammals; similar conclusions were reached on the basis of limited species cross-reactivity of antibodies (reviewed by Wiche, 1989; and unpublished data).

The tissue-specific expression of plectin mRNA was examined by hybridizing total RNA isolated from various tissues of rat, including heart, leg muscle, brain (5- and 60-d old), liver and kidney, to clone C1 cDNA. In all the tissues examined, the probe detected a single RNA band of ~15 kb (Fig. 6). The hybridization signals varied considerably, being highest in skeletal muscle and lowest in liver.

Evidence for A Single Copy Gene in Rat

Genomic Southern blot analysis was carried out to obtain information about the number of plectin genes in the rat genome. Genomic DNA from rat glioma C_6 cells was digested to completion with the restriction enzymes XbaI, EcoRI, HindIII, and BstEII. After electrophoresis in agarose gels, DNA fragments were blotted and hybridized to the 4,300-bp insert of clone C1, which lacks recognition sites for the enzymes used. As shown in Fig. 7, only single restriction fragments of various sizes were detected by the probe in all of the digests. This suggested that rat plectin is encoded by a single-copy gene.

cDNA Sequence and Predicted Structural Characteristics of Plectin

Rat cDNA clones were sequenced according to Sanger et al. (1977), using a variety of sequencing strategies (see above). The total length of the plectin clones sequenced was 14,682 bp. A potential initiation codon (position 1) was preceded by termination codons in all three reading frames (TGAs and TAG at positions -14, -43, -116, and -243), and was followed by a continuous open reading frame of 12,420 nucleotides encoding 4,140 amino acids (Fig. 8). Because the NH_2 terminus of plectin was blocked (J. S. Vandekerckhove, unpublished data), we were unable to confirm the NH_2-terminal sequence by protein sequencing. The coding region was followed by 1,164 bp of 3′ untranslated sequence. The polyadenylation signal was found 21 bp upstream of the polyA tail. A map showing the sites of relevant restriction enzymes is presented in Fig. 1.

The expected molecular weight of the protein encoded by the 12,420 nucleotides contained within the open reading frame was 466,481. Its predicted isoelectric point was 6.32. The deduced polypeptide chain was composed of 33% hydrophobic amino acids (alanine, leucine, isoleucine, phenylalanine, tryptophane, and valine) and of 31.4% charged residues (aspartate, glutamate, lysine, and arginine). With a few exceptions (see below) most amino acids were distributed equally throughout the protein. Hydropathy analysis of the sequence according to Kyte and Doolittle (1982) revealed

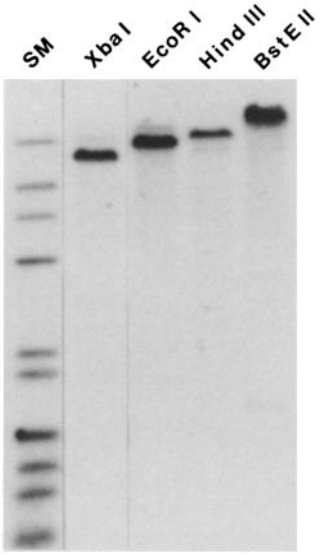
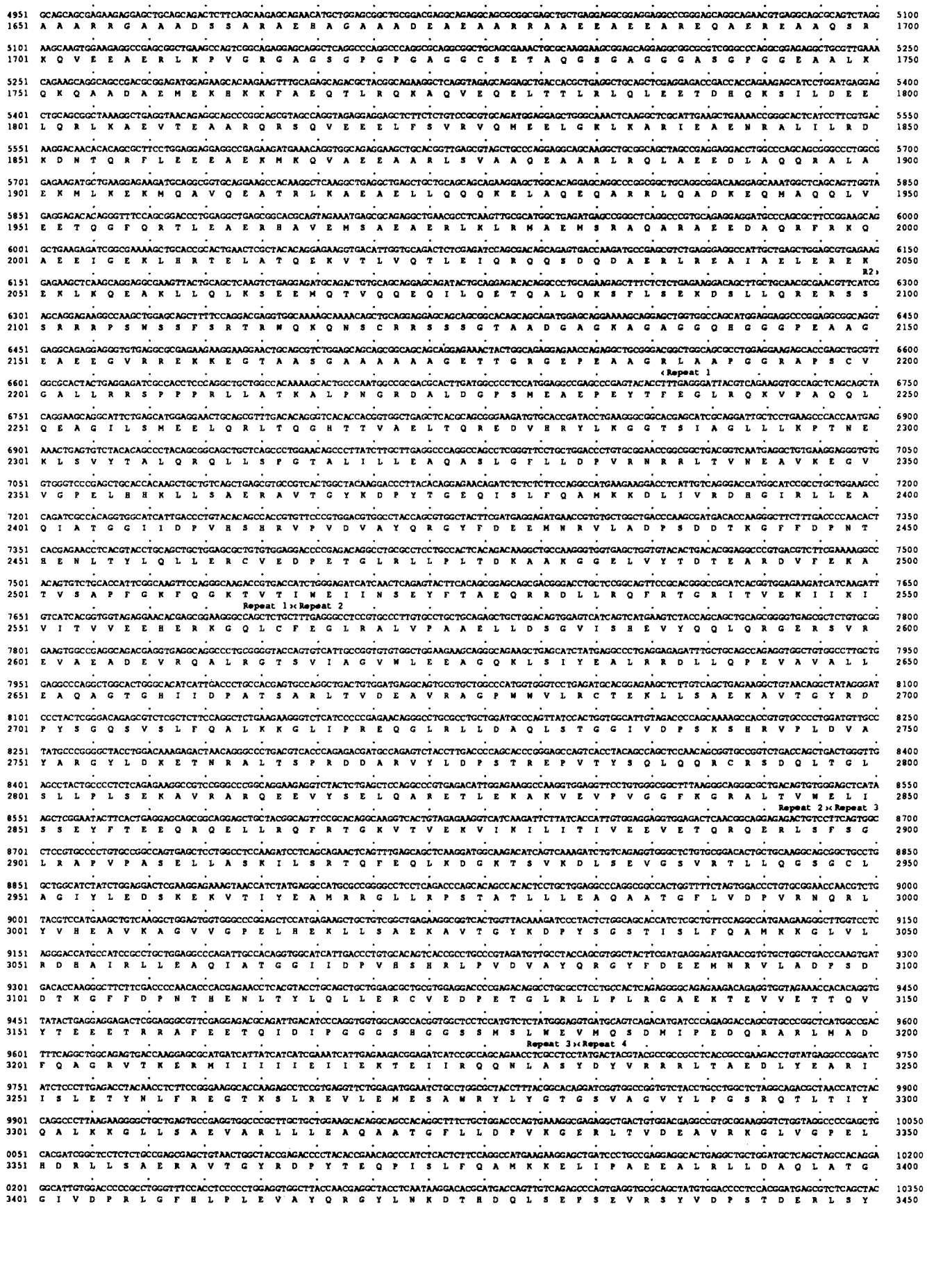


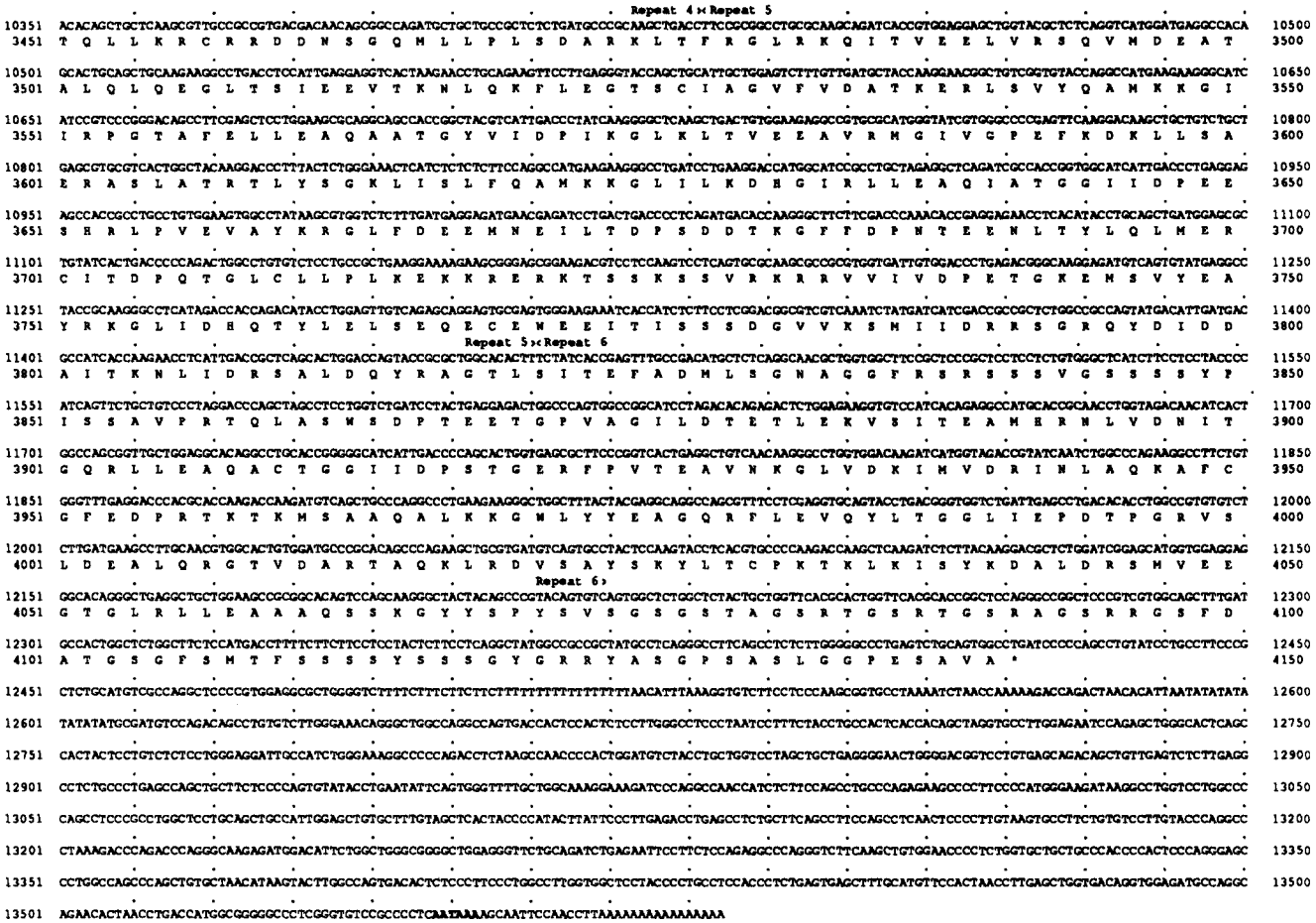
Figure 7. Southern blot analysis of genomic rat DNA. Aliquots (3 μg) of rat glioma C_6 cell genomic DNA were digested with the indicated restriction enzymes and probed with the insert of clone C1. *SM*, fragments of HindIII-cut lambda DNA and HaeIII-cut ØX174 DNA corresponding to 23.1, 9.4, 6.6, 4.4, 2.3, 2.0, 1.4, 1.1, 0.9, and 0.6 kb.

Figure 8. Nucleotide and derived amino acid sequence of rat plectin. Nucleotide and amino acid residues are numbered relative to the putative initiation codon. Stop codons in all three reading frames upstream of this site are underlined. Rod domains R1 and R2 and each of the six tandem repeat domains in the COOH-terminal region are marked (<>). The polyadenylation signal found at the end of the transcript is highlighted. These sequence data are available from the EMBL data library under accession number X59601.

B



C



several significant hydrophobic regions. The first (residues 436–452) corresponded to a potential transmembrane domain, if analyzed according to the method of Klein et al. (1985); however, other methods failed to confirm this prediction. A second hydrophobic stretch of nucleotides, separated from the first by a long hydrophilic domain, was located near the middle of the molecule, beginning at residue 2,168.

Electron micrographs of plectin indicate that it is a highly elongated molecule, with a long central rod domain and globular domains at each end (Foisner and Wiche, 1987). Inspection of the predicted amino acid sequence showed a long stretch of residues that had the heptad repeat characteristic of alpha-helical coiled coils (Crick, 1953; McLachlan and Stewart, 1975) so that in successive groups of seven residues, a-b-c-d-e-f-g, residues *a* and *d* were usually hydrophobic or apolar (Fig. 9). Fourier analyses of this region of the sequence showed a very strong periodicity at 7/2 residues (see Fig. 12 *B*) confirming the presence of the heptad repeat. The probability of such a strong periodicity occurring by chance was 10^{-9} (Table I). Moreover, secondary structure predictions based on the method of Garnier et al. (1978) indicated that this region of the sequence had an extremely high potential for the alpha conformation. This heptad-rich region extended from approximately residue 700 to 2,100 and clearly corresponded to the central rod domain observed by electron microscopy (Fig. 10). The two flanking regions (residues 1–700 and 2,100–4,140) did not have such a high potential for alpha structure, had frequent proline residues (which would be expected to interrupt the formation of extensive alpha helices) and did not have a distinct heptad pattern. We therefore identify these regions with the globular domains observed by electron microscopy. The precise boundary between the NH_2-terminal globular domain and the rod was not completely clear. The heptad structure of the sequence was very pronounced after residue 930 but before that was punctuated by a series of extensive regions in which the heptad motif was weak or absent and where there were often proline and other residues typical of less regular secondary structure. This effect became more pronounced as one moved towards the NH_2 terminus and so the boundary drawn at approximately residue 700 is somewhat arbitrary, although it is close to the NH_2-terminal end of a reasonably long heptad-rich region. The precise location of the start of the rod domain does not significantly influence our subsequent analysis of the sequence data. Overall, therefore, we interpret the derived amino acid sequence in terms of the model illustrated in Fig. 10, in which there is a central rod domain (extending from residues 700 to 2,100) flanked by globular domains at the NH_2 (domain GN) and COOH (domain GC) termini. A more comprehensive picture of the

fgabcde	efgabcd	
738	930	
PVFNTLR	PKVQSGS	
DELRGAQ	ESVIQEY	
EVGERLQ	VDLRTRY	
QRHGERD	SELTTLT	
VEVERWR	SQYIKFI	
ERVTQLL	SETLRRM	
ERWQAVL	EEEERLA	
AQTDVRQ	EQQRAEE	
RELEQLG	RERLAEV	
RQLRYYR	EAALEKQ	
ESAD	RQLAEAH	
	AQAKAQA	
	ELEAREL	
abcdefg	QRRMQEE	*
	VTRREEA	
831	AVDAQQQ	
PIANSQA	KRSIQEE	
AREQLRQ	LQHLRQS	
EKALLEE	SEAEIQA	*
IERHGEK	KAQQVEA	
VEECQKF	AERSRMR	
AKQYINA	IEEEIRV	
IKDYELQ	VRLQLET	
LITYKAQ	TERQRGG	
LE	AEDELQA	
	LRARAEE	
	AEAQKRQ	
	AQEEAER	
	LRRQVQD	
	ESQRKRQ	
	AEAELAL	
	RVKAEAE	
	AAREKQR	
	ALQALDE	
	LKLQAEE	
	AERRA	

Figure 9. Regions rich in heptad repeats between residues 700 and 1,200. There are extensive regions between 738–812 and 831–888 (shown on the left-hand side), but these are much less extensive than the regions in rod subdomain R2, such as A (930–1,140) shown on the right. The mainly hydrophobic residues that form the heptad repeat are shown in bold type. Note that the phase of the heptad changes in region A (*).

likely structure and function of each domain emerged from a detailed analysis using a range of different computer-based methods.

Rod Domain

Although the rod domain extended from approximately residue 700 to 2,100, its detailed character changed along its length. The regular heptad repeat was punctuated several times by small regions which probably disrupted the coiled coil conformation locally. Such regions may represent areas of local flexibility that may have a role in the assembly of the protein or its interaction with other cytoskeletal components. The frequency of these interruptions was, however, greater in the NH_2-terminal region of the rod between roughly residues 700 and 900 as illustrated in Fig. 9. Dot matrix analysis (Fig. 11) also indicated that the rod tended to divide into two regions: R1, from residue 700 to 930, in which there was little evidence of internal repeats; and R2, from residue 930 to 2,100, where there was evidence of a highly repetitive structure that produced a "box" of short lines of homology similar to that seen, for example, in the rod domain of myosin. Moreover, it was clear that there was a substructure to this box so that there were lighter stripes (*arrows* in Fig. 11) roughly every 200 residues that indicated that there were short stretches in which the repetitive motif was absent. A similar effect was seen, for example, with the dystrophin sequence (Cross et al., 1990), and these light stripes indicated that the R2 region was composed of five subregions (A, B, C, D, and E in Figs. 10 and 11) that probably reflect the generation of this portion of the plectin sequence by a series of gene duplication events. The approximate boundaries of these subregions were 930–1,140 (*A*), 1,180–1,380 (*B*), 1,420–1,580 (*C*), 1,750–1,940 (*D*), and 1,940–2,100 (*E*). There may also be a shorter helical segment between residues 1,670 and 1,720. The short stretches or "linkers" between A, B, C, and D contained proline and other residues often associated with turns in polypeptide chains and so probably represented regions of local flexibility similar to those seen in region R1 of the rod. Overall, subregions A, B, and C tended to form one group and regions D and E a second. There was not a very clear boundary between D and E and so these could form a continuous coil about 360-residues long. The linker between C and D was remarkably basic (calculated pI = 12.4 compared with 5.3 and 5.1 for the flanking regions) and, in addition, contained a cluster of five cysteine residues that may be important for intra- as well as intermolecular crosslinks.

Table I. Fourier Data for Regions A, B, and C of Rod Segment R2 (Residues 930–1,580)

Residue type (residues)	Period	Scaled Fourier Intensity*	Probability of arising by chance*
Acidic	10.4	23.2	8×10^{-11}
Basic	10.4	16.9	5×10^{-8}
Hydrophobic	7/2	20.7	10^{-9}

* Scaling and probability calculation as described by McLachlan and Stewart (1976).

Fourier analysis of region R2 showed the presence of a very strong repeat in the charged amino acids with a frequency corresponding to 10.4 residues (Fig. 12 *A*). Both acidic and basic residues showed this peak which was extremely unlikely to have arisen by chance (Table I). The basic and acidic periodicities were, however, out of phase, so that the sequence in the R2 region of the rod was made up from a series of zones of alternating positive and negative charge. This highly repetitive pattern in the sequence clearly corresponded to the short repetitive pattern detected by dot-matrix analysis (Fig. 11). A similar pattern in the charged residues with a frequency of 10.4 residues has been observed in the desmoplakin sequence (Green et al., 1990), whereas myosin has a repeat at 28 residues and intermediate filaments at 28/3 residues (reviewed by Stewart et al., 1989). By analogy with other fibrous proteins such as myosin, tropomyosin, and intermediate filaments, such a periodic variation in charge is probably involved in an association between plectin

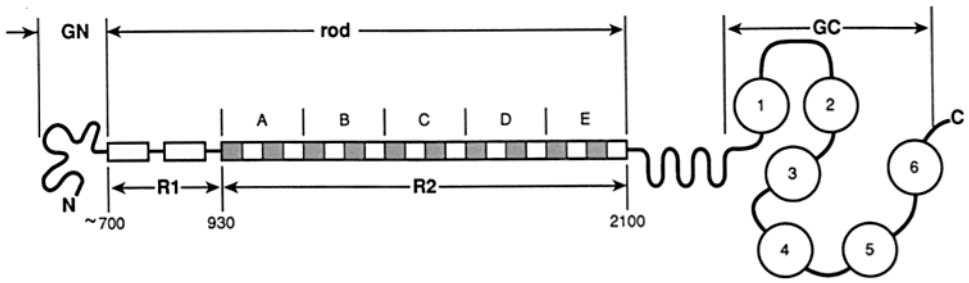


Figure 10. Highly schematic representation of the model for the macromolecular structure of plectin proposed on the basis of EM and analysis of the derived amino acid sequence. The sequence divides naturally into three overall domains: a globular NH_2-terminal domain (*GN*); a central fibrous domain (*rod*) spanning roughly residues 700–2,100, that probably has an alpha-helical coiled coil conformation; and a globular COOH-terminal domain (*GC*). The rod domain can be further subdivided into a long R2 region, in which there is a strong 10.4-residue periodicity in the charged residues, and a shorter R1 region where short alpha-helical stretches are joined by linkers that appear to have a less regular conformation. It is not clear where the precise boundary between GN and R1 is located, but it appears to be close to residue 700. The R2 region appears to be constructed from five tandem repeats (A, B, C, D, and E) of ~200 residues that are separated by short linkers of variable length, whereas the GC domain has a remarkably strongly conserved sixfold repeat about 335 residues long that probably has a more globular conformation. Only a single polypeptide chain is illustrated in this diagram, but the presence of the characteristic heptad repeat in hydrophobic residues over most of the rod region indicates that the plectin molecule is almost certainly constructed from two chains that form an alpha-helical coiled coil. As discussed in detail in the text, the arrangement of the linkers between the stretches of heptad-containing sequence indicates very strongly that the two chains in the molecule are arranged parallel and in register.

molecules staggered so that zones of opposite charge would complement one another (see Stewart et al., 1989).

A number of properties of the amino acids sequence in the rod region indicated that the two chains in the plectin coiled coil are probably arranged parallel to one another and in register. Coiled coils form by interdigitating the hydrophobic residues that form a stripe along one side of each helix. Clearly this "knobs-in-holes" arrangement (Crick, 1953) can be formed when the two chains are either parallel or antiparallel, but the influence of the hydrophobic effect (because of the exclusion of water from the aliphatic side chains of the hydrophobic residues in the heptad) is greatest when the overlap between the two alpha helices is complete. The correctness of this deduction is confirmed by the observation that in coiled coils such as myosin (Stewart, 1982), tropomyosin (Stewart, 1975), and intermediate filament proteins (Quinlan and Franke, 1982), the two chains in the coil are in register. In the case of plectin, the coil-coil in the rod domain appears to have a number of short linker regions along it, and because these linker regions are spaced aperiodically along the sequence, it is clear that the greatest degree of overlap between the helical segments will be obtained when the two chains are arranged parallel and in register. In this arrangement, the linker regions in one helix will be opposite those in the other and so there would be complete overlap between all helical segments. If the chains were to be ar-

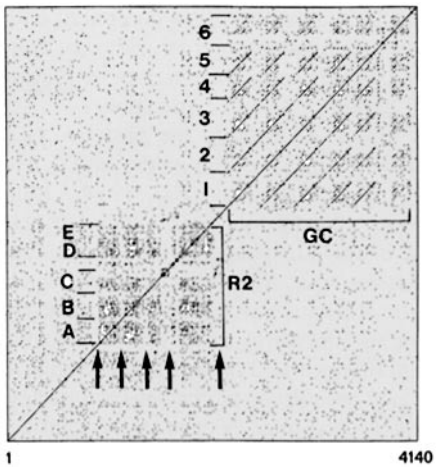


Figure 11. Dot matrix self comparison of the plectin sequence. A boxlike area in the center of the sequence results from the highly repetitive R2 region of the rod. There are narrow lighter bands (*arrows*) separating this region into five subdomains A, B, C, D, and E. There is also a very strong sixfold tandem repeat in the COOH-terminal globular domain GC.

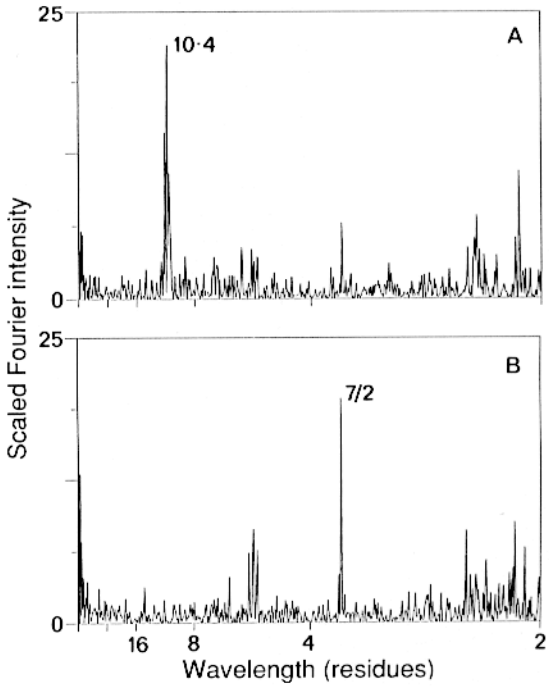


Figure 12. Fourier transforms of regions A, B, and C of rod segment R2 (residues 930–1,940). *A*, acidic residues; *B*, hydrophobic residues. There are very strong peaks corresponding to a periodicity of 10.4 residues in the charged residues (data for basic residues not shown) and to 7/2 in the hydrophobic residues.

Table II. Amino Acid Conservation in Plectin's Six Repeat Domains

Repeat	1st 2,238–2,565	2nd 2,566–2,897	3rd 2,898–3,232	4th 3,233–3,477	5th 3,478–3,820	6th 3,821–4,073
			Percent Identity			
1st 2,238–2,565		54.6	61.6	49.4	47.0	30.9
2nd 2,566–2,897	73.6		50.1	51.8	41.4	34.2
3rd 2,898–3,232	75.8	69.6		49.0	49.1	35.0
4th 3,233–3,477	70.6	71.0	74.3		44.9	29.6
5th 3,478–3,820	68.0	61.4	68.9	70.0		28.9
6th 3,821–4,073	51.8	54.2	54.0	51.9	51.0	
			Percent Homology*			

* Conservation of identical plus similar residues. Sequences were aligned using the program GAP (UWGCG). The score is based on the number of identical or similar residues and the length of the consensus segment. Conservation of identical plus similar residues.

ranged antiparallel or out of register (either parallel or antiparallel), then it would not be possible to place all the linkers opposite one another and so complete matching between the helical regions could not be obtained. Similarly, the 10.4-residue charge repeat present in region R2 would be most effective when the chains were arranged parallel and in register as indeed has been found in all other fibrous proteins that have analogous charge repeat patterns.

COOH-Terminal Globular Domain

The third major domain of plectin (GC) extended from residue 2,238 to residue 4,073, near the COOH terminus of the protein. The GC domain was separated from the R2 rod domain by a very basic region (pI calculated as 10.0) of 137 residues, 10 of which were proline. A striking characteristic of the GC domain was the presence of six remarkably long tandem repeats (Fig. 11). Each repeat contained 335 ± 8 residues, except for repeats 4 and 6 which contained 245 and 253, respectively. The repeat boundaries are shown in Fig. 8. Dot-matrix comparison and alignment of the repeats showed that their sequences were highly homologous. As shown in Table II, repeats 1–5 showed 41–62% identity and 61–76% homology over their entire lengths, with repeats 1, 3, and 5 being most strongly related (Table II; and unpublished results obtained with the program ALIGN, DNA Star Inc.). Repeat 6 was the most distant (29–35% identity; 51–54% homology). These long repeats probably reflect a series of gene duplication events in the generation of the plectin sequence.

All six long GC repeats shared a strongly conserved central region that was constructed from nine tandem repeats of a 19-residue motif plus two flanking partial repeats as illustrated in Fig. 13. These 19-residue motifs frequently had a proline residue located at their COOH terminus and a possible consensus sequence for the motif would be:

E T G E R L L E A Q A A T R G L/Y L/V D P

The most conserved residue within the 19-amino acid motif was at position 15, which was occupied predominantly by glycine (46 out of 60). In the last repeat, however, this residue was changed to cysteine, with a single exception.

The 19-residue periodicity was also detected in Fourier transforms of the GC region (Fig. 14). Green et al. (1990) detected a 9.5-residue periodicity in the charged residues of desmoplakin and we found a similar periodicity in the charged residues of plectin. Like Green et al. (1990) we found this periodicity was extremely strong and the probability that such a periodicity could have arisen by chance was 3×10^{-12} for the acidic residues and 2×10^{-8} for the basic. However, we think it is likely that this periodicity is an overtone (19/2) of the underlying 19-residue periodicity. The strong series of peaks at orders of 19 residues in the proline Fourier transform (Fig. 14), for example, indicates that this is the fundamental periodicity and that the peak at 19/2 for the charged residues comes about because the 19-residue motif contains two equally-spaced clusters of charged residues within it. Similarly we found no Fourier evidence for a periodicity at 38 residues (2×19) as had been proposed for desmoplakin (Green et al., 1990) and none of the overtone peaks we observed in any Fourier transform indexed on this longer spacing.

Fourier transforms of the potential for different secondary structures for the GC region only showed strong peaks for the beta conformation (data not shown), probably mainly because of the groups of hydrophobic residues (with high beta potential) that characterized this region (Fig. 13). The comparatively high number of glycine and aspartic acid residues, together with the frequent proline residues in this region suggest that it would probably not have a high content of alpha secondary structure and there was no evidence for the heptad repeats characteristic of interacting alpha helices. Overall we envisage that this region would most likely fold to form an essentially globular structure in which the chain would weave back and forth with the proline-rich regions associated with surface turns in the direction of the polypeptide chain and most of the hydrophobic residues located in the interior, perhaps frequently in short stretches having a predominantly beta conformation. Such a structure would be analogous to that found in the immunoglobulin domain fold,

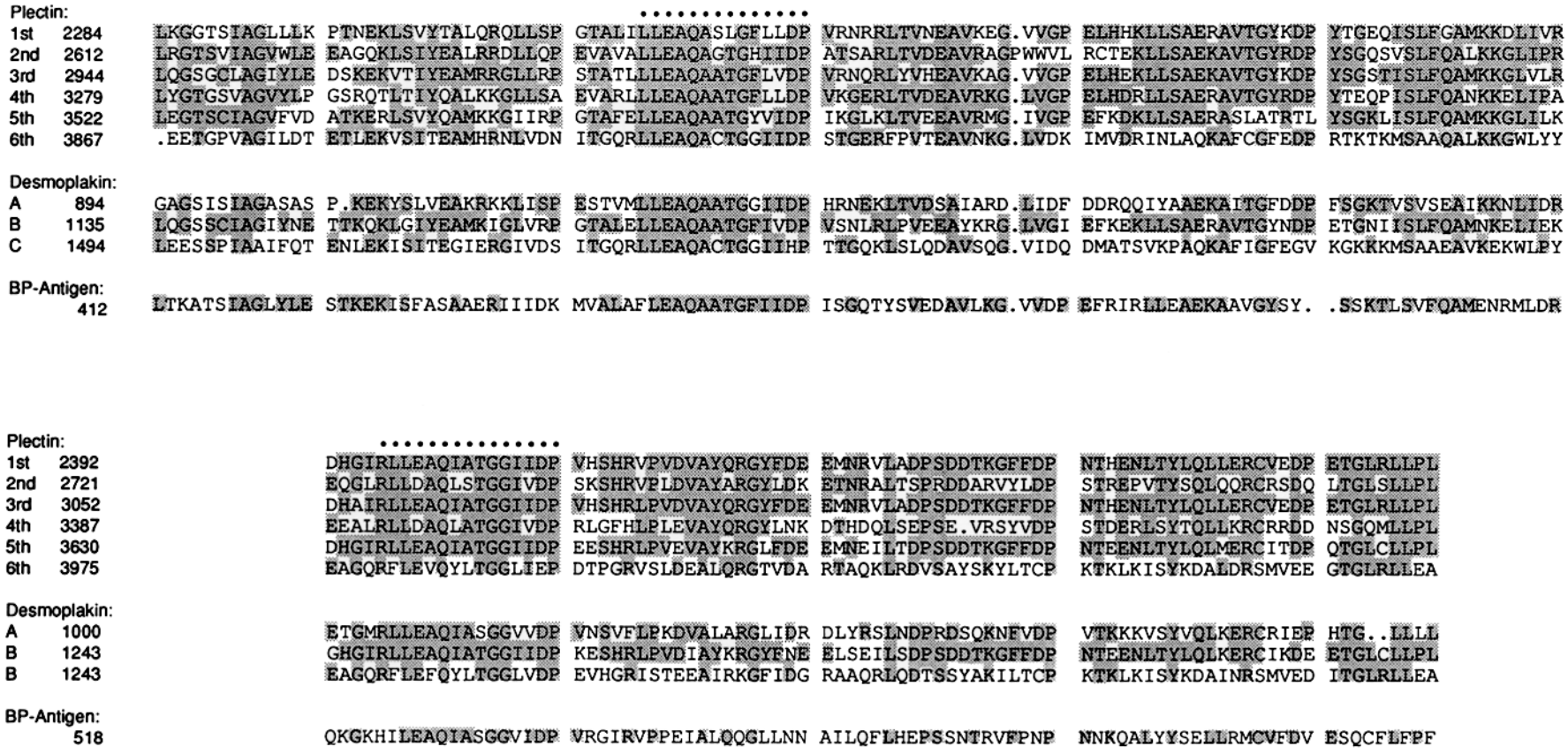


Figure 13. Sequence comparison of 19-amino acid repeat motif common to rat plectin, human desmoplakin, and bullous pemphigoid (BP)-antigen. Alignment of each of the six COOH-terminal domains of plectin with themselves, with desmoplakin's COOH-terminal A, B, and C domains and with the sequence of BP-antigen was done using the program GAP (UWGCG) and the alignment was then slightly refined by eye; sequence gaps thereby created are indicated by dots. The sequences shown are restricted to the regions containing the repeat motifs. Numbers in second column on the left-hand side indicate positions of first residues shown in each line, according to Fig. 8 (*plectin*), Green et al. (1990) (*desmoplakin*), and Stanley et al. (1988) (*BP-antigen*). The most conserved residues are highlighted in grey, and two stretches of 14 and 15 identical residues in the 2nd and 6th repeat motifs are marked by full circles. Note, that the 19-amino acid motif is repeated consecutively nine times in full in all sequences, with two partial repeats of 8 and 9 residues at the 5′ and 3′ ends, respectively.

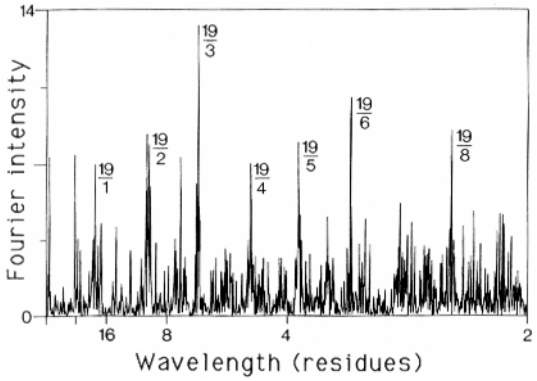


Figure 14. Fourier transform of the proline residues of repeats 1–4 of the GC COOH-terminal globular domain sequence. Note the strong series of peaks that index as orders of a fundamental repeat of 19 residues.

but we stress that we do not wish to imply any direct structural homology between the plectin GC region and immunoglobulins. Although the periodicity of 9.5 residues is close to that observed in the charged residues of intermediate filament proteins (McLachlan and Stewart, 1982; Conway and Parry, 1988), we do not find most of the conformational data strongly supportive of the notion that the GC region contains extensive areas having a coiled coil structure so that the plectin and intermediate filament charge repeats may complement one another as has been proposed for the analogous repeats in desmoplakin (Green et al., 1990).

The last 67 residues of the sequence (4,074–4,140), immediately following the sixth long GC repeat were highly basic (calculated pI = 11.4) like the region bridging between the rod and the first GC repeat. The last 67 residues also contained a GSRX motif (where X is largely hydrophobic) repeated four times, followed by a cluster of serine residues.

Plectin has been shown to be a prominent target for a number of protein kinases both in vivo and in vitro (Herrmann and Wiche, 1983; 1987). In an examination of the amino acid sequence for phosphorylation recognition motifs consensus sites for the following kinases were found: (*a*) cAMP-dependent protein kinase (RRXS; Feramisco et al., 1980), three times; (*b*) protein kinase C (RXXS/TXR; Kemp and Pearson, 1990), three times; (*c*) Ca^{2+}/calmodulin kinases (RXXS/T; Kemp and Pearson, 1990; Blackshear et al., 1988), 35 times; and (*d*) cdc2 cell cycle kinase (S/TPXZ; Moreno and Nurse, 1990), two times. The majority of these potential recognition sites was located towards the end of the rod, and towards the end of the GC domain. The functional significance of these sites remains to be shown. The sites along the rod may be important for mechanisms controlling plectin's self association and/or its interactions with other coiled coil structures, such as the rod domains of intermediate filaments proteins (Foisner et al., 1988; 1991*a*). Moreover, phosphorylation of plectin by cdc2 kinase, as has already been demonstrated in vitro (unpublished results), might be associated with the cytoskeletal reorganization occurring at mitosis.

Homology to Desmoplakin and Bullous Pemphigoid Antigen

A comparison of the cDNA and protein sequence of plectin with other sequences available from data bases (SWISSPROT and EMBL) revealed a striking homology of plectin's COOH-terminal repeat domains to COOH-terminal repeat domains of human desmoplakin (Green et al., 1990). Using a sequence-matching protein alignment computer program (GAP; UWGCG) plectin's 1st, 3rd, and 5th repeats were found to share 49.2, 51.4, and 64.4% of identical amino acids, respectively, with domain B of desmoplakin. Repeat 6 showed 58.2% identity with domain C of desmoplakin, and the extreme carboxyl termini of both proteins were 67.4% identical (Table III). These values were as high as the ones obtained when plectin's six repeats were compared among themselves

Table III. Comparisons between Repeat Domains of Plectin, Desmoplakin, and BP Antigen

	Desmoplakin*				
	Domain A 849–1,090	Domain B 1,091–1,433	Domain C 1,434–1,704	COOH-terminus 1,705–1,753	BP Antigen‡ 368–644
Plectin					
1st 2,238–2,565	42.6	49.2	26.5		37.0
2nd 2,566–2,897	44.6	40.1	30.0		32.8
3rd 2,898–3,232	42.6	51.4	29.7		36.4
4th 3,233–3,477	38.2	46.9	22.0		36.3
5th 3,478–3,820	42.2	64.4	28.3		40.4
6th 3,821–4,073	31.8	33.3	58.2		26.0
COOH-terminus 4,074–4,140				67.4	

Scores (% identity) were obtained as in Table II.
* Note that starts of domains A–C of desmoplakin as defined here, do not correspond to those given in Green et al. (1990).
‡ Stanley et al. (1988).

(Table II), indicating a remarkably high degree of conservation. The sequence alignment of plectin and desmoplakin in the repeats' most conserved region is shown in Fig. 13. The highest scores of identical residues were obtained for two stretches of 15 and 14 amino acids found in the 6th and 2nd 19-amino acid repeat motifs, respectively (Fig. 13). The striking conservation of glycines at position 15 of the repeat motifs of plectin, including their conversion to cysteine in the last full repeat, was also observed in desmoplakin.

Other domains of the plectin and desmoplakin sequences showed no significant sequence homologies, not even when subjected to dot matrix analysis at very low stringency (window size: 20; stringency: 6). It should be noted, however, that the 10-residue periodicity of charged amino acids found in plectin's rod section (see above) matches a similar one observed in desmoplakin's rod (Green et al., 1990). In line with the reported homology between desmoplakin domains A, B, and C and bullous pemphigoid (BP) antigen (Green et al., 1990; Stanley et al., 1988), plectin's six repeat domains were homologous to the last 276 amino acids of BP antigen (Fig. 13, last line; and Table III). Overall, the level of conservation between plectin and desmoplakin was higher than that between desmoplakin and BP antigen (Green et al., 1990) or between plectin and BP antigen.

Discussion

Plectin is a versatile cytoplasmic cross-linking protein, whose interaction partners range from constituent proteins of the subplasma membrane skeleton and the nuclear lamina to intermediate filament proteins and microtubule-associated proteins. In addition, the protein self-associates giving rise to networklike arrays. As part of a comprehensive investigation of plectin's properties and role in the organization of the cytoskeleton, we have isolated and sequenced the cDNA encoding the rat protein species. This information has enabled us to make predictions regarding the structure of the protein and to tentatively map a few of its functional epitopes. Questions related to the tissue expression of plectin transcripts and their conservation between different species were also addressed. The information obtained will serve as a basis for a precise mapping of structural epitopes involved in plectin's diverse interactions. Furthermore, it will facilitate the investigation of the cellular role of this protein using such tools as in vitro mutagenesis and transfection of cells.

The authenticity of the clones was established by several criteria: (*a*) All the clones detected an apparently single high molecular weight mRNA species of the expected size (~15 kb) on Northern blots (Fig. 4; and data not shown); (*b*) affinity-purified antibodies to plectin, showing no cross-reactivity with any cellular proteins other than plectin (Fig. 2), specifically recognized the proteins encoded by the different clones; and, most convincingly, (*c*) cDNA-encoded polypeptides were immunoreactive with a series of different well-characterized mAbs to plectin, as demonstrated by Western blotting.

All mAb epitopes examined were mapped to regions in the NH_2-terminal half of the polypeptide chain, which contained the sequence encoding the coiled coil rod segment. This was fully consistent with ultrastructural studies showing that these mAbs decorated the rod domain of plectin molecules, on the one hand, and with the location of antigenic sites as predicted by computer analysis of plectin's amino acid sequence (Jameson and Wolf, 1988). In fact, the highest point of the obtained antigenicity profile coincided with the region corresponding to residues 1,600–1,800 (i.e., the overlapping segments of clones C1 and C2) where the epitopes of half of the mAbs examined were encoded (Fig. 6).

Plectin transcripts were found in all tissues examined, albeit at different levels. The relatively high level of plectin mRNA found in brain was somewhat unexpected, considering that plectin has not been detected in neuronal cell processes by immunofluorescence microscopy (Wiche, 1989). Thus, it is likely that the mRNA in brain tissue originated mostly from neuroglia. This would be consistent with the significantly decreased levels of plectin mRNA found in young brain, where the ratio of glial to neuronal cells is considerably lower than in adult brain (Ling and Leblond, 1973).

The calculated molecular weight (466,481) of the protein encoded by the open reading frame sequence exceeded by far the molecular weight of plectin deduced by SDS-PAGE (~300,000). This discrepancy was probably due to the known aberrant migration of fibrous proteins subjected to SDS-PAGE. Such behavior would be consistent with our observation that a plectin fusion protein encoded by a continuous 8.2-kb-long cDNA sequence in SDS-PAGE exhibited an apparent molecular weight of only 270,000 instead of the expected 320,000 (unpublished data). A precedent of this kind would be dynein heavy chain. This protein comigrates with plectin in SDS-PAGE (Koszka et al., 1987; and data unpublished), but its actual molecular weight estimated from EM and the size of specific UV-cleavage products is >400,000 (Lye et al., 1987; Vallee et al., 1988). Alternatively, the size difference could be explained if the protein as isolated from cells and tissues does not represent the whole polypeptide synthesized, but only part of it. The smaller protein isolated could then represent a proteolytically derived form of a larger precursor (processed in vivo or degraded during isolation). This possibility seems not unlikely, in view of preliminary results (unpublished data) that show that an anti-plectin positive protein of apparent molecular weight considerably >300,000 is indeed present in extracts of cultured cells. However, it remains to be established whether this protein is a genuine precursor form of the 300K species or represents a different, though at least in part structurally related, species.

The ultrastructural data and the model derived from them can be reconciled with the plectin sequence data in several ways. The likely extent of the central rod domain was between 1,000 and 1,400 residues (after making allowance for the nonhelical linker regions), which would correspond to an alpha-helical coiled coil between 150- and 210-nm long. Thus the roughly 190-nm-long central rod section of dumbbell-shaped plectin structures observed by rotary shadowing EM (Foisner and Wiche, 1987) appeared to be in reasonable agreement with the predicted length of the rod domain. It was not completely clear, however, how many chains constituted the particles seen by EM. The globular domains observed in the shadowed particles were roughly equal in size, whereas the two globular domains identified from analysis of the amino acid sequence were probably rather different, with the COOH-terminal domain being substantially larger than the NH_2-terminal. Although there was some uncertainty

as to the precise boundary between the GN domain and the rod, it seemed improbable, even with the most distal boundary, that this domain would be of comparable size to the GC domain. Therefore, if the particles observed by EM contained two intact plectin polypeptide chains, it would seem most likely that the chains were arranged antiparallel. However, the characteristics of the rod sequence and particularly the arrangement of small linker regions argued strongly that the chains were more likely to be arranged parallel, as found for all other fibrous alpha-helical coiled coil cytoskeletal proteins investigated. A dimeric arrangement of chains would still appear feasible, however, if plectin was proteolysed posttranscriptionally or during isolation, as would not be unexpected, given the discrepancy between the polypeptide molecular weight predicted from the sequence and that observed by SDS-PAGE (see above). In this case, in the shadowed molecules either the size of the GC domain could have been reduced (so that the two globular domains were of equal size) or the GN domain and part of the rod were removed so that the particles observed were constructed from two molecules joined near their NH_2 termini. However, we feel that it was probably more likely that the particles observed by EM contained four complete chains arranged into two molecules. The chains in each molecule would be parallel and the two molecules would be arranged antiparallel and overlap by almost their entire length so that the globular regions observed at each end of the "dumbbell" by EM would be constructed from the GN domains of one molecule and the GC domains of the other. Such tetrameric assemblies conceivably could also be formed by partially proteolysed molecules lacking end domains. Clearly it is important to resolve the precise structure of these particles observed by shadowing, which may be possible now that polypeptides corresponding to distinct cDNA domains can be expressed and purified, and their behavior studied in vitro. However, regardless of the precise structure of these particles, it is clear that there is no difficulty in reconciling their appearance with the three-domain model derived from the sequence data.

As demonstrated by EM of mAb-decorated plectin structures, a number of mAbs with epitopes residing within a 20–25-nm stretch, close to the center of the rod, inhibited plectin–vimentin and plectin–lamin B interactions in vitro (Foisner et al., 1991*a*). Provided that the rod segments of the antibody-decorated structures corresponded in length to the predicted rod sequence of the cDNA (Fig. 10), the epitopes of these inhibitory mAbs can roughly be mapped to residues 1,180–1,350. This corresponded to a segment which partially overlapped with both areas identified as coding regions for the epitopes of these and a number of other mAbs (Fig. 2). This segment therefore is likely to play a role in plectin's interactions with vimentin and lamin B. In light of recent data showing that both of these interactions were affected by protein kinase A as well as protein kinase C phosphorylation (Foisner et al., 1991*b*), it may be significant that this region is flanked by sequences rich in serine and threonine residues, although typical recognition consensus sequences for these kinases were not identified there.

The structural similarity found between rat plectin and human desmoplakin is intriguing. Both proteins contain large segments of tandemly repeated 19-residue motifs, arranged within multiple repeat domains at their COOH termini. However, while only three such repeats, referred to as A, B, and C domains (Green et al., 1990), are found in human desmoplakin, plectin contained twice as many. Considering their lengths, the repeat domains of plectin showed a striking structural conservation among themselves, although clearly they were not identical. The homology between the various repeat domains of plectin and those of desmoplakin varied between 22 and 64% identical residues (49–80% similar residues). Noteworthy, the B domain of desmoplakin was more closely related to repeats 1, 3, and 5 of plectin than to any of the other desmoplakin repeats (Table II; and Green et al., 1990). The structural conservation of these repeats may suggest that these domains are involved in similar functions of both proteins. It has been proposed that the repeats of desmoplakin may constitute contact sites for intermediate filament proteins (Green et al., 1990), although no biochemical evidence for such function of desmoplakin has yet been forthcoming. Plectin, on the other hand, has been demonstrated to indeed interact with a variety of intermediate filament proteins (Herrmann and Wiche, 1987; Foisner et al., 1988). However, some of the epitopes involved at least in the interactions of plectin with vimentin and lamin B have been mapped to plectin's rod domain, using mAb inhibition assays (Foisner et al., 1991*a*). Although this does not rule out a participation of the COOH-terminal plectin repeats in these interactions, it makes it less likely. Furthermore, it seems unlikely that the GC domain has an extensive coiled coil structure (which would be necessary to match the periodicities in each molecule) and so the 9.5-residue periodicity, which seems to reflect the 19-residue motif structure, may only be fortuitously the same as that observed in intermediate filament proteins.

No significant sequence homologies were detectable using computer search programs, in parts other than the COOH-terminal repeat domains of both proteins. However, their rod domains displayed almost identical periodicities of charged residues. These periodicities are quite likely to play important roles in the lateral association of the rod segments with molecules of their own kind or with different protein species of similar characteristics. Thus, considering that plectin and desmoplakin molecules have similar ultrastructure, albeit distinct dimensions, desmoplakin may be seen as a smaller version of plectin. The functions of these proteins must be quite different, however. Desmoplakin has been reported to occur specifically at desmosomal plaques in epithelial cells, while plectin was found in many cell types and at many distinct locations. Moreover, in contrast to desmoplakin, a variety of proteins have been identified that specifically interact with plectin at the molecular level. Nevertheless, it will be a challenging task to clarify whether these and other structurally related proteins, such as BP antigen (Stanley et al., 1981), which has been suggested to play an important role in basal cell–substrate interaction, share common functions.

In conclusion therefore, we have established the primary structure of plectin by sequencing cDNA corresponding to the rat protein and have shown that plectin mRNA is present in a range of tissues, with greatest abundance in muscle. Analysis of the derived amino acid sequence is consistent with a three-domain molecular structure based on a central alpha-helical coiled coil and globular NH_2- and COOH-terminal domains. There are extensive repeating motifs within both the rod and GC domain, some of which probably

derive from gene duplication events in the generation of the plectin sequence. Some of these repeats, particularly that at 10.4 residues in the R2 rod segment, are probably involved in the assembly of plectin molecules into larger aggregates. The availability of these sequence data and the structural models derived from them will enable the structure and interactions of plectin to be mapped more precisely using such methods as directed mutagenesis and expression in both prokaryote and eukaryote systems. Eventually such analysis may even give some clues regarding the still enigmatic biological role of intermediate filaments, one of plectin's major interacting partners.

We thank Heidi Kroiss, University of Vienna, for her valuable assistance in sequencing, and Christian Fritze, University of Chicago, for help with different computer programs. We would also like to thank Dr. Elaine Fuchs, HAMI-University of Chicago, for access to DNA software programs available in her laboratory.

This work was supported by grants from the Austrian Science Research Fund (Österreichischer Fonds zur Förderung der wissenchaftlichen Forschung) and the Wüscheck-Dreher Foundation (W.-D. Stiftung).

Received for publication 13 February 1991 and in revised form 5 April 1991.

References

Aviv, H., and P. Leder. 1972. Purification of biologically active globin mRNA by chromatography on oligothymidylic acid-cellulose. *Proc. Natl. Acad. Sci. USA.* 69:1408–1412.

Blackshear, P. J., A. C. Nairn, and J. F. Kuo. 1988. Protein kinases 1988: a current perspective. *FASEB (Fed. Am. Soc. Exp. Biol.) J.* 2:2957–2969.

Chou, P. Y., and G. D. Fasman. 1978. Empirical predictions of protein conformation. *Annu. Rev. Biochem.* 47:251–276.

Conway, J. F., and D. A. D. Parry. 1988. Intermediate filament structure: 3. Analysis of sequence homologies. *Jut. J. Biol. Macromol.* 10:79–98.

Crick, F. H. C. 1953. The packing of alpha-helices: simple coiled-coils. *Acta. Crystallogr.* 6:689–697.

Cross, R. A., M. Stewart, and J. Kendrick-Jones. 1990. Structural predictions for the central rod domain of dystrophin. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 262:87–92.

Dayhoff, M. O. 1979. Atlas of protein sequence and structure. Vol. 5, National Biomedical Research Foundation, Washington, D.C.

Emini, E. A., J. V. Hughes, D. S. Perlow, and J. Boger. 1985. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J. Virol.* 55:836–839.

Feinberg, A. P., and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132:6–13.

Feramisco, J. R., D. B. Glass, and E. G. Krebs. 1980. Optimal spatial requirements for the location of basic residues in peptide substrates for the cyclic-AMP dependent protein kinase. *J. Biol. Chem.* 255:4240–4245.

Foisner, R., and G. Wiche. 1987. Structure and hydrodynamic properties of plectin molecules. *J. Mol. Biol.* 198:515–531.

Foisner, R., and G. Wiche. 1991. Intermediate filament-associated proteins. *Curr. Opin. Cell Biol.* 3:75–81.

Foisner, R., F. E. Leichtfried, H. Herrmann, J. V. Small, D. Lawson, and G. Wiche. 1988. Cytoskeleton-associated plectin: in situ localization, in vitro reconstitution, and binding to immobilized intermediate filament proteins. *J. Cell Biol.* 106:723–733.

Foisner, R., B. Feldman, L. Sander, and G. Wiche. 1991*a*. Monoclonal antibody mapping of structural and functional plectin epitopes. *J. Cell Biol.* 112:397–405.

Foisner, R., P. Traub, and G. Wiche. 1991*b*. Protein kinase A and C-regulated interaction of plectin with lamin B and vimentin. *Proc. Natl. Acad. Sci. USA.* 88:3812–3816.

Garnier, J. D., J. Osguthorpe, and B. Robson. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* 120:97–120.

Green, K. J., D. A. D. Parry, P. M. Steinert, M. L. A. Virata, R. M. Wagner, B. D. Angst, and L. A. Nilles. 1990. Structure of human desmoplakins: implications for function in the desmosomal plaque. *J. Biol. Chem.* 265: 2603–2612.

Gubler, U., and B. J. Hoffman. 1983. A simple and very efficient method for generating cDNA libraries. *Gene.* 25:263–269.

Han, J. H., C. Stratowa, and W. J. Rutter. 1987. Isolation of full length putative rat lysophospholipase cDNA using improved methods for mRNA isolation and cDNA cloning. *Biochemistry.* 26:1617–1625.

Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene.* 28:157–162.

Hermann, H., and G. Wiche. 1983. Specific in situ phosphorylation of plectin in detergent-resistant cytoskeletons from cultured chinese hamster ovary cells. *J. Biol. Chem.* 258:14610–14618.

Herrmann, H., and G. Wiche. 1987. Plectin and IFAP 300K are homologous proteins binding to microtubule associated proteins 1 and 2 and to the 240 kilodalton subunit of spectrin. *J. Biol. Chem.* 262:1320–1325.

Huynh, T. V., R. A. Young, and R. W. Davis. 1985. Constructing and screening cDNA libraries in lambda gt10 and in lambda gt11. *In* DNA cloning: a practical approach. Vol. 1. D. M. Glover, editor. IRL Press Limited, Oxford, England 49–78.

Jameson, B. A., and H. Wolf. 1988. The antigenic index: a novel algorithm for predicting antigenic determinants. *Comput. Appl. Biosci.* 4:181–186.

Kemp, B. E., and R. B. Pearson. 1990. Protein kinase recognition sequence motifs. *Trends Biochem. Sci.* 15:342–346.

Klein, P., M. Kanehisa, and C. DeLisi. 1985. The detection and classification of membrane spanning proteins. *Biochim. Biophys. Acta.* 815:468–476.

Koszka, C., R. Foisner, H. M. Seyfert, and G. Wiche. 1987. Isolation of a Ca^{2+}-protease resistant high M_r microtubule binding protein from mammalian brain: characterization of properties partially expected for a dynein-like molecule. *Protoplasma.* 138:54–61.

Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157:105–132.

Ling, E.-A., and C. P. Leblond. 1973. Investigation of glial cells in semithin sections. *J. Comp. Neur.* 149:73–82.

Lipman, D. J., and W. R. Pearson. 1985. Rapid and sensitive protein similarity searches. *Science (Wash. DC).* 227:1435–1441.

Lye, R. J., M. E. Porter, J. M. Scholey, and J. R. McIntosh. 1987. Identification of a microtubule-based cytoplasmic motor in the nematode C. elegans. *Cell.* 51:309–318.

Maizel, J., and R. Lenk. 1981. Enhanced graphic matrix analysis of nucleic acid and protein sequences. *Proc. Natl. Acad. Sci. USA.* 78:7665–7669.

Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 545 pp.

Maurer-Fogy, I., C. P. M. Reutelingsperger, J. Pieters, G. Bodo, C. Stratowa, and R. Hauptmann. 1988. Cloning and expression of cDNA for human vascular anticoagulant, a Ca^{2+}-dependent phospholipid-binding protein. *Eur. J. Biochem.* 174:585–592.

McLachlan, A. D., and M. Stewart. 1975. Tropomyosin coiled-coil interactions. *J. Mol. Biol.* 97:293–304.

McLachlan, A. D., and M. Stewart. 1976. The 14-fold periodicity in tropomyosin and the interaction with actin. *J. Mol. Biol.* 103:271–298.

McLachlan, A. D., and M. Stewart. 1982. Periodic features of intermediate filament protein sequences. *J. Mol. Biol.* 162:693–698.

Messing, J. 1983. New M13 vectors for cloning. *Methods Enzymol.* 101:20–78.

Moreno, S., and P. Nurse. 1990. Substrates for $p34^{cdc2}$: In vivo veritas? *Cell.* 61:549–551.

Noegel, A. A., S. Rapp, F. Lottspeich, M. Schleicher, and M. Stewart. 1989. The Dictyostelium gelation factor shares a putative actin binding site with alpha-actinins and dystrophin and also has a rod domain that contains six 100-residue motifs that appear to have a cross-beta conformation. *J. Cell Biol.* 109:607–618.

Pearson, W. R., and D. J. Lipman. 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA.* 85:2444–2448.

Quinlan, R. A., and W. W. Franke. 1982. Heteropolymer filaments of vimentin and desmin in vascular smooth muscle and cultured BHK cells demonstrated by chemical crosslinking. *Proc. Natl. Acad. Sci. USA.* 79:3452–3456.

Rasmussen, U. B., P. Basset, and J.-Y. Daniel. 1989. Direct amplification of cDNA inserts from λ libraries using the cloning-adaptor as primer for PCR. *Nucleic Acids Res.* 17:3308.

Reed, K. C., and D. A. Mann. 1985. Rapid transfer of DNA from agarose gels to nylon membranes. *Nucleic Acids Res.* 13:7207–7221.

Robson, B., and E. Suzuki. 1976. Conformations of amino acid residues in globular proteins. *J. Mol. Biol.* 107:327–356.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463–5467.

Simon, M.-N., R. Mutzel, H. Mutzel, and M. Vernon. 1988. Vectors for expression of truncated coding sequences in Escherichia coli. *Plasmid.* 19:94–102.

Staolen, R. 1982. Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. *Nucleic Acids Res.* 10: 4731–4751.

Stanley, J. R., P. Hawley-Nelson, S. H. Yuspa, E. M. Shevach, and S. I. Katz. 1981. Characterization of bullous pemphigoid antigen: a unique basement membrane protein of stratified squamous epithelia. *Cell.* 24:897–903.

Stanley, J. R., T. Tanaka, S. Mueller, V. Klaus-Kovtun, and D. Roop. 1988. Isolation of complementary DNA for bullous pemphigoid antigen by use of patients' autoantibodies. *J. Clin. Invest.* 82:1864–1870.

Stewart, M. 1975. Tropomyosin: no stagger between chains. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 53:5–7.

Stewart, M. 1982. Chain register in myosin rod. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 140:210–212.

Stewart, M. 1990. Intermediate filaments: structure, assembly and molecular interactions. *Curr. Opin. Cell Biol.* 2:91–100.

Stewart, M., R. A. Quinlan, R. D. Moir, S. R. Clarke, and S. J. Atkinson. 1989. The role of repeating sequence motifs in interactions between alpha-helical coiled-coils such as myosin, tropomyosin and intermediate filament proteins. *In* Cytoskeletal and Extracellular Proteins. U. Aebi and J. Engel, editors. Springer Verlag, Heidelberg, Germany. 150–159.

Vallee, R. B., J. S. Wall, B. M. Paschal, and H. S. Shpetner. 1988. Microtubule-associated protein 1C from brain is a two-headed cytosolic dynein. *Nature (Lond.)*. 332:561–563.

Vieira, J., and J. Messing. 1982. The pUC plasmids, and M13 mp7-derived systems for insertion mutagenesis and sequencing with universal primers. *Gene.* 19:259–268.

Wiche, G. 1989. Plectin: general overview and appraisal of its potential role as a subunit protein of the cytomatrix. *CRC Crit. Rev. Biochem.* 24:41–67.

Wiche, G., and M. A. Baker. 1982. Cytoplasmic network arrays demonstrated by immunolocalization using antibodies to a high molecular weight protein present in cytoskeleton preparations from cultured cells. *Exp. Cell Res.* 138:15–29.

Zinn, K., D. DiMaio, and T. Maniatis. 1983. Identification of two distinct regulatory regions adjacent to the human β-Interferon gene. *Cell.* 34: 865–879.