An alternatively spliced region of the human hexabrachion contains a repeat of potential N-glycosylation sites

(type III homology units/epidermal growth factor motifs/fibrinogen/primary structure)

JEFFREY R. GULCHER, DONALD E. NIES, LINDA S. MARTON, AND KARI STEFANSSON*

Departments of Neurology and Pathology (Neuropathology), and the Committees on Neurobiology and Immunology, University of Chicago, Chicago, IL 60637

ABSTRACT   We have cloned and sequenced two cDNA molecules that code for parts of two forms of human hexabrachion. The smaller clone has a sequence that corresponds to the previously published sequence of a cDNA clone coding for a part of chicken hexabrachion [Jones, F. S., Burgoon, M. P., Hoffman, S., Crossin, K. L., Cunningham, B. A. & Edelman, G. M. (1988) Proc. Natl. Acad. Sci. USA 85, 2186–2190]. It has eight consecutive domains similar to the type III homology units from fibronectin, several epidermal growth factor repeats, and a domain similar to the β and γ chains of fibrinogen. The larger clone has 5' and 3' ends that are identical to the smaller clone but also has an alternatively spliced 1.9-kilobase internal segment. The unique segment contains remarkable repeats of potential glycosylation sites and an additional seven type III homology units.

We have previously described a pair of large glycosylated polypeptides that we isolated from human fetal brain and called gp150/225 (1). Both polypeptides form homomultimers linked by disulfide bonds, both contain a significant amount of N-linked oligosaccharides, and both have the HNK-1 epitope that may be a marker for adhesion molecules (2). Recently it has become clear that one and the same protein or a family of closely related proteins has been described under several names, including hexabrachion (3), tenascin (4), myotendinous antigen (5), glioma mesenchymal extracellular matrix antigen (6), J1 glycoprotein (2), and cytotactin (7). The gp150/225 also belongs to this family (see below). Rotary shadowed electron micrographs of this protein(s) reveal a characteristic structure, which lead Erickson and Inglesias (3) to come up with the name hexabrachion. The typical molecule consists of a pair of three arms that are connected through a central globular particle (8). In this paper we have, for clarity and convenience, elected to use the name hexabrachion for all of the above-mentioned proteins. Although it has not been unequivocally proven that they are all the same protein, the circumstantial evidence is compelling (9). The name hexabrachion describes the typical structure without implying anything about function or localization, both of which are still somewhat controversial.

Most reports on the hexabrachions indicate that they are polypeptides of 200 kDa or more, but there are also reports on smaller polypeptides in the family, including our own on the gp150 (1). It has, however, been debated whether or not these smaller polypeptides are authentic hexabrachions (9, 10). We believe that some of that data presented below support the opinion that they are.

Recently, Jones et al. (11) reported on the sequence of a cDNA clone (λC801) coding for a part of one of the hexabrachions from the chicken. The deduced amino acid sequence contained several epidermal growth factor (EGF)

repeats, fibronectin type III homology units, and a domain with a sequence similar to the β and γ chains of fibrinogen. Here we report on the sequence of two cDNA clones coding for a part of two forms of human hexabrachion. One clone (P31) is equivalent to λC801, and P31 and λC801 have sequences that are 80% similar to each other. The other clone has in addition a 1.9-kilobase (kb) internal segment with remarkable repeats of potential N-glycosylation sites and seven type III homology units. During the final stages of the preparation of this paper, a report by Pearson et al. (12) was published describing sequences of two cDNA clones (cTn8 and cTn10) coding for a part of the chicken hexabrachion. These clones overlap with the 5' end of λC801, and cTn10 reaches beyond it and codes for a total of 13 EGF repeats.

METHODS

RNA Isolation and Generation of a cDNA Library. Total RNA from the human glioblastoma line U-373MG (ATCC no. HTB-17) was isolated by using the guanidinium isothiocyanate/cesium chloride method (13). The preparation was enriched in mRNA by using an oligo(dT)-cellulose affinity column (14). A λgtll cDNA library was made by using Gubler and Hoffman's modification of Okayama and Berg's method (15). Lysate-absorbed antiserum against hexabrachion was used to probe plate replicas (16). Positive clones were plaque purified and subcloned into Bluescript (Stratagene) for mapping and sequencing.

DNA Sequencing. Nested deletions of subclones chosen for sequencing were made with exonuclease III (17). Single-stranded phage were induced, and the DNA was sequenced by using dideoxynucleotide termination (18) with Sequenase (United States Biochemical) (19). A 265-base-pair (bp) segment containing the intrinsic EcoRI site of P29 was amplified from P29 with the polymerase chain reaction by using flanking oligonucleotide primers (20). The amplified segment was sequenced through the EcoRI site to determine whether the EcoRI fragment subclones of P29 are contiguous. Both strands of all subclones were sequenced and analyzed with the sequence analysis software package of the University of Wisconsin Genetics Computer Group (21). Peptide sequences were compared with the Dayhoff table and the GAP program in the above software package.

Northern, Southern, and Western Blotting. Total RNA (8 μg per lane) from the human glioblastoma cell line U87MG (ATCC no. HTB-14) was electrophoresed on a 1% agarose/formaldehyde gel and capillary blotted onto a nylon membrane. Southern blotting of human lymphocyte DNA was carried out as described using a nylon membrane (22). cDNA fragments were eluted from an agarose gel and labeled by random priming (23). The blots were washed with 0.1× SSC (1× SSC = 0.15 M NaCl/0.015 M sodium citrate, pH 7) at

---

Abbreviation: EGF, epidermal growth factor.
*To whom reprints should be addressed at: Department of Neurology, University of Chicago, 5841 South Maryland Avenue, Chicago, IL 60637.

Developmental Biology: Gulcher *et al.*

Proc. Natl. Acad. Sci. USA 86 (1989) 1589

60°C twice for 30 min and exposed with intensifiers at −70°C. Proteins were electrophoresed on a SDS/PAGE gel (24), electroblotted to nitrocellulose (25), and immunostained with polyclonal antibodies that had been affinity purified on the polypeptide backbone of hexabrachions isolated from human brain as described (1).

RESULTS

cDNA Clones from a Human Glioblastoma Cell Line. Eight positive clones from a glioblastoma cell line that produces hexabrachions were plaque purified, and mapping revealed that they fell into two sets of overlapping clones. Fig. 1 shows the restriction maps of the longest and representative clone of each set, P29, which is 5 kb, and P31, which is 3 kb. Inspection of the corresponding maps shows that both the 3' and 5' ends of the clones contain identical restriction sites and that they differ apparently only in an internal 1.9-kb segment found in P29 but not in P31.

Evidence for Alternative Splicing. Total RNA from the human glioblastoma cell line U87MG was electrophoresed, blotted, and probed with two cDNA probes. One probe was the 700-bp *Bam*HI fragment from P29, which is common to both cDNA clones. The other probe was the internal 1.2-kb *Sac* II–*Eco*RI fragment of P29, which is within the segment unique to P29 and is not present in P31. As Fig. 2A shows, the common probe demonstrated messages of 6.2 and 7.8 kb. The unique probe only bound to the 7.8-kb message.

A Western blot in Fig. 2B of the electrophoresed proteins from U87MG cell proteins stained with antiserum to the polypeptide backbone of the hexabrachions shows two polypeptides with apparent molecular masses of 240 and 180 kDa.

The Southern blot in Fig. 2C suggests there is a single gene. Human genomic DNA, when cut with three different restriction enzymes and probed with the small *Bam*HI fragment of P29, which is also in P31, revealed a single restriction fragment in each digest. In addition, a gene dosage study with a genomic clone suggests there is one copy of the gene per haploid genome (J.R.G. and K.S., unpublished results). The 1.9-kb internal region that is found in P29 and not in P31 is of sufficient size to account for the difference between the larger (240 kDa) and the smaller (180 kDa) protein. The similarities of P29 and P31, the correlation between the messages and the proteins, and the fact that both appear to be derived from one gene suggest that they may be derived from alternate splicing of one primary message.

Sequence. Fig. 3 shows the nucleotide sequence and deduced amino acid sequence of clones P29 and P31. P29 has 4877 bases and P31 has 2658 bases. The sequences of the overlaps at the 3' and 5' ends are identical. P29 extends an additional 329 bases past the 5' end of P31, but the 3' ends are even. The internal region unique to P29 is 1890 bases long. Both clones are one continuous open reading frame. The N terminus of the sequence deduced from P29 contains six
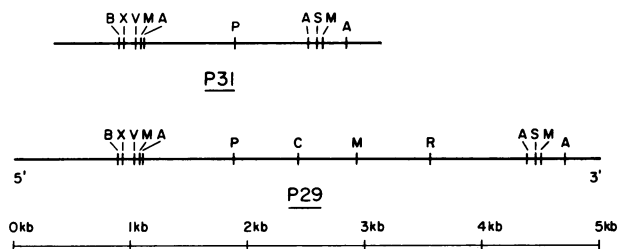

FIG. 1. Restriction maps of P29 and P31, which are the largest clones of each of the two groups of clones. The 5' ends of both clones contain six identical restriction sites and the 3' ends contain four identical restriction sites. However, P29 has three restriction sites in a unique internal segment not found in P31. A, *Apa* I; B, *Bam*HI; C, *Sac* II; M, *Sma* I; P, *Pst* I; R, *Eco*RI; S, *Sst* I; V, *Eco*RV.
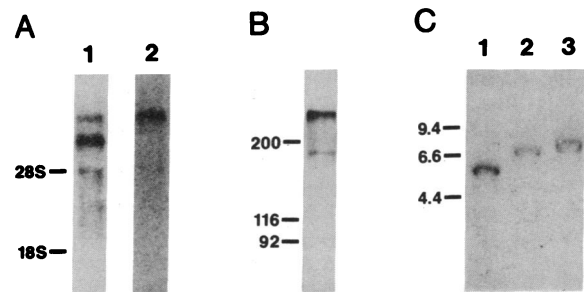

FIG. 2. (A) Both lanes of this Northern blot contain 8 μg of electrophoresed total RNA from the human glioblastoma cell line U87MG. Lane 1 was probed with the 700-bp *Bam*HI–*Eco*RI fragment of P29, which is shared by P31. Prominent bands at 6.2 kb and 7.8 kb are seen along with nonspecific binding of 28S rRNA. Lane 2 was probed with the 1.2-kb *Sac* II–*Eco*RI fragment of P29, which is not found in P31. Only the 7.8-kb band is bound by the probe. (B) Western blot of electrophoresed proteins of U87MG glioblastoma cells harvested at the same time as the RNA in A. The blot was immunostained with antiserum to hexabrachion, demonstrating polypeptides of 180 and 240 kDa. (C) Southern blot containing human lymphocyte DNA cut with *Eco*RI (lane 1), *Bam*HI (lane 2), and *Pst* I (lane 3). The blot was probed with the 700-bp *Bam*HI fragment of P29, which is shared by P31. Only one restriction fragment in each digest hybridized to the probe. This suggests that P29 and P31 are derived from one gene.

repeats of 31 amino acids that have six cysteine residues in exactly the same places in each repeat (Fig. 4A). The repeats consist of $X_4CX_3CX_5CX_4CXCX_8C$ (where X is an amino acid). Repeats of this kind have been described in several molecules and are often called EGF motifs. If conservative substitutions are considered, each repeat in the P29 derivative has a sequence that is 80% similar to the sequence of any of the others.

Immediately 3' to the EGF motifs in the P29 derivative (Fig. 5), there are 15 (repeats III 1–15) domains similar to the type III homology units of fibronectin, and there are 8 repeats in the P31 derivative (repeats III 1–5 and repeats III 13–15) (Fig. 4B). Apart from repeats III 6–9, each homology unit is equally similar to such units in fibronectin as it is to other type III repeats in hexabrachion. The repeat III 3 has an RGD sequence (Fig. 3). Repeat III 6 begins exactly where the derivative of the segment unique to P29 begins. The repeats III 6–9 are remarkable for ordered placement of potential N-glycosylation sites. The potential glycosylation sites can be regarded as a pair of segments containing NLT-X_{11}-NWT-X_{23}-NLT-X_{62}-NWT-X_{23}-NLT or four segments of NLT-X_{11}-NWT-X_{23}-NLT, where the second and the fourth repeats are missing the first NLT sequence. Repeat III 10 also has three potential N-glycosylation sites, but they are NVS-X_8-NLS-X_{27}-NIS. Hence the spacing of the N-glycosylation sites is off, and the sites use serine rather than threonine as the third residue. The repeats III 6–9 show 80% sequence similarity to each other. However, they are no more or less similar to any of the other type III homology units derived from P29 than they are to the type III homology units in fibronectin (calculations not shown). Repeat III 12 ends exactly where the derivative of the unique internal segment of P29 ends. Repeat III 15 is followed by an 8-amino acid segment, the sequence of which in our analysis did not appear to be strikingly similar to any known sequence. However, the C-terminal-most domain of 68 amino acids has a sequence that is equally similar to the sequences of the β and γ chains of fibrinogen as they are to each other.

DISCUSSION

In this report we describe the cloning and sequencing of two cDNA clones that code for a portion of two forms of human hexabrachion. The difference between P29 and P31 is a

1890-nucleotide internal segment that is unique to P29; and 3′ and the 5′ ends are the same. P31 corresponds to a clone (λC801) from a chicken cDNA library previously reported by Jones *et al.* (11). The sequence similarity of P29 and λC801 is 80% if conservative substitutions are considered. P29 and P31 are derived from a single gene. Thus the difference between them is apparently due to alternative splicing of the primary message transcribed from the hexabrachion gene.

There are two polypeptides secreted by the glioblastoma cell line U87MG that, on immunoblots, react with antibodies against the polypeptide backbones of the hexabrachions. They have apparent molecular masses of 240 and 180 kDa,
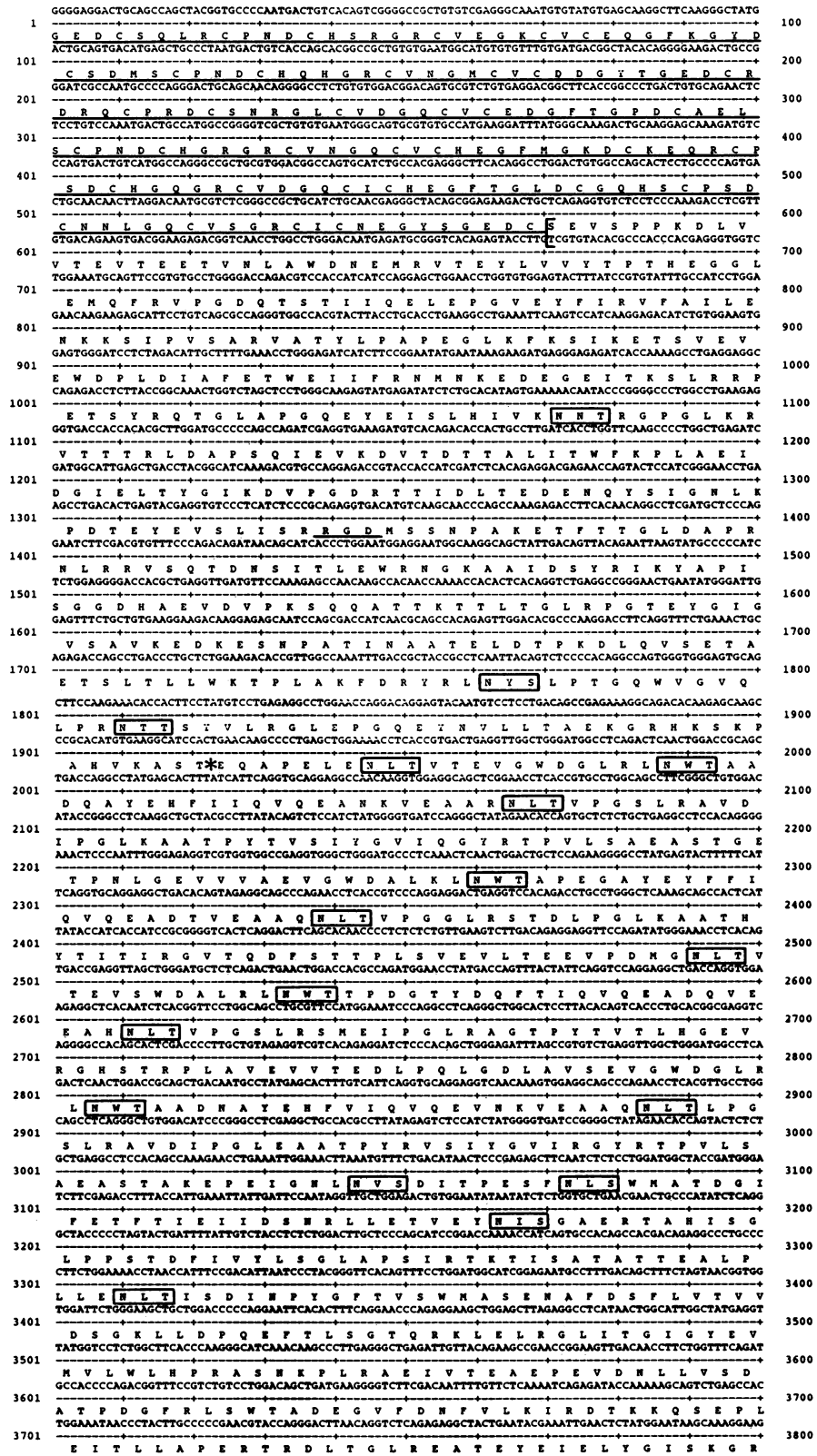

FIG. 3. *(Figure continues on top of opposite page.)*

Developmental Biology: Gulcher *et al.*
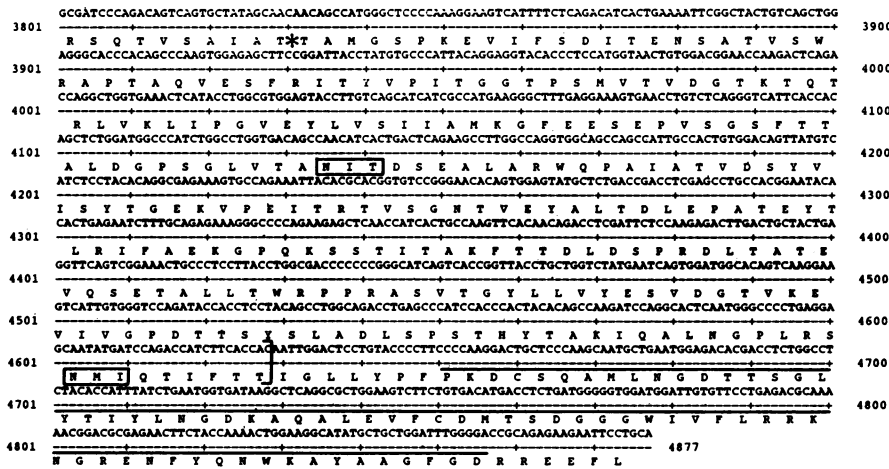
Proc. Natl. Acad. Sci. USA 86 (1989)   1591

FIG. 3.   Sequence of P29 and below it the deduced amino acid sequence. The stars mark the two ends of the internal segment unique to P29. The potential N-glycosylation sites are boxed. The beginning of the first type III homology unit and end of the last one are demarcated by brackets, the cysteine-rich region and the RGD sequence are underlined, and the domain similar to fibrinogen is overlined.

which are similar to the sizes seen in human fetal brains (225 and 150 kDa, respectively). A probe common to P29 and P31 hybridizes to two messages on a Northern blot containing RNA from U87MG cells. A probe for the region unique to P29 does not hybridize to the smaller message. This indicates that P31 is probably derived from the smaller message, which most likely codes for the 180-kDa form. The internal region unique to P29 is large enough (1.9 kb) to account for the 60-kDa difference in apparent molecular masses of the two forms of the protein. These data also indicate that the smaller protein is derived from the same gene as the larger protein and is therefore an authentic hexabrachion.

Both P29 and P31 have extensive sequence similarities to fibronectin and in particular to the type III homology units of fibronectin. The type III homology units of fibronectin contain both the RGD sequence (26), which binds to integrins, as well as an additional cell-binding domain (27). They also contain a high-affinity heparin-binding site and a DNA-binding domain (28). Both P29 and P31 have an RGD sequence in repeat III 3. It is an RRGDM rather than a classic GRGDS sequence, and it is still unclear whether it grants the protein sufficient affinity for cells to allow us to classify it as a cell-binding site. It is well to keep in mind that although the RGD sequence has been

found in several adhesion proteins it is also present in more than 120 proteins that probably have nothing to do with adhesion (27). However, the RGD sequence in the hexabrachions is located in the same place within a type III homology unit as is the RGD sequence in fibronectin.

Each of the type III homology units in the hexabrachion is as similar to the type III units in fibronectin as it is to other type III homology units in the hexabrachion, with the exception of repeats III 6–9, which are remarkable for containing the systematically placed potential glycosylation sites. However, each of the units in repeats III 6–9 is equally similar to the fibronectin type III homology units as it is to the type III units in the hexabrachion, apart from the other three in repeats III 6–9 . The similarity in the type III units between these two molecules may indicate that they both acquired the type III units, through exon shuffling and subsequent reduplication, at the same time or that each homology unit in both molecules has independently diverged as much from an ancestral protein as it can without losing the ability to perform its function. Repeats III 6–9 are the homology units that contain the remarkably rigidly organized potential glycosylation sites, and they have sequences that are 80% similar to each other. Repeats III 6 and III 8 each contains three potential glycosy-
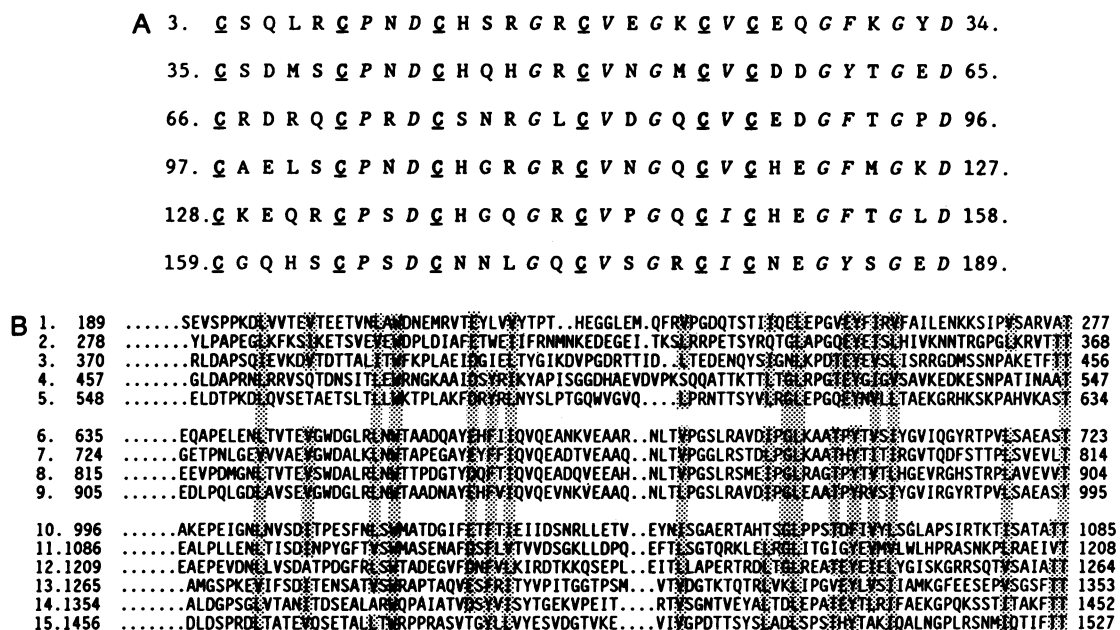

FIG. 4.   (A) The six cysteine-rich repeats at the N terminus of the deduced sequence of P29 are shown. The cysteines are underlined. (B) Type III homology units from the derivative of P29. The units are numbered 1–15. Repeats III 6–12 are unique to the P29 derivative. Some of the sequence similarities are shaded.
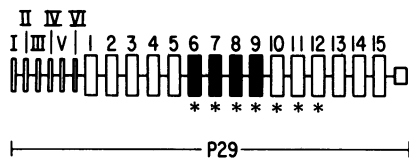
FIG. 5. Schematic drawing of the P29 derivative. First there are six cysteine-rich repeats marked by Roman numerals I–VI. They are followed by 15 type III homology units marked by Arabic numerals 1–15. The type III units encoded for by the segment unique to P29 are marked by asterisks. Repeats III 6–9 containing the ordered potential N-glycosylation sites are blackened. The small box at the end represents the domain similar to fibrinogen.

lation sites, and each is followed by a homology unit with two potential glycosylation sites that correspond exactly to the second and third sites in repeats III 6 and III 8. Either repeats III 6–9 have evolved through recent reduplication or there has been substantial evolutionary pressure to maintain a highly ordered structure. Where that evolutionary pressure would come from is not clear, but it may be of relevance that the hexabrachions contain the HNK-1 epitope, which resides on an N-linked oligosaccharide (1). The HNK-1 moiety, which is on several other molecules including neural cell adhesion molecule, neuronal–glial cell adhesion molecule, and myelin-associated glycoprotein, has been shown to mediate adhesion between cells (29). One possibility is that the potential glycosylation sites on repeats III 6–9 contain the HNK-1 carbohydrate and through that mediate a biologically important adhesion among cells or between cells and extracellular matrix. Even more interesting is that repeats III 6–9 are contained within the alternatively spliced region defined by our clones. Hence there are forms of the protein that do and forms that do not contain this domain, and this may be important for regulation of cell–cell interactions. The HNK-1 epitope is not likely to be confined to this domain because it is also on the smaller forms that are probably coded for by messages that do not contain the unique region, but alternate splicing may modulate adhesion by changing the number of HNK-1 carbohydrates.

The homology units III 10–12 also belong to the unique portion of P29. Repeat III 10 contains three glycosylation sites as does repeats III 6–9, but they are NVS, NLS, and NIS rather than NLT, NWT, and NLT, and their spacing is different (NVS-X_8-NLS-X_{27}-NIS). Repeat III 11 only has one glycosylation site. It is of interest that repeat III 10 has diverged only slightly from repeats III 6–9 as far as the nature and placement of potential glycosylation sites is concerned, but it is not in a class with the repeats III 6–9 when it comes to sequence similarity. In addition, this region in the hexabrachion is a good example of how reduplicated exons (or groups of exons) may evolve into units that, although maintaining overall similarity, contain somewhat different sequences dictating different functions. It is also of interest here that contactin (30) and the L1 glycoprotein (31), both of which may mediate intercellular interactions in the nervous system, contain two and three type III repeats, respectively.

At the N terminus of the polypeptide sequences deduced from both P29 and P31 is a cysteine-rich domain that consists of 6 repeats of $X_4CX_3CX_5CX_4CXCX_8C$ in P29 and 2.5 in P31. These repeats are similar to cysteine-rich repeats found in the EGF precursor, the EGF receptor (32), the low density lipoprotein receptor (33), and many other proteins.

At the C terminus of the polypeptides deduced from both P29 and P31 is a 68-amino acid domain with a sequence that is as similar to the β and γ chains of human fibrinogen as they are to each other. These regions in the β and γ chains of fibrinogen contain a loop that is formed by an intrachain disulfide bond between two cysteines (34) also present in the

hexabrachion sequence. We are tempted to postulate that there is such a loop in the corresponding part of the hexabrachion. It has been reported that fibronectin binds to the hexabrachions (10). This binding may occur through the fibrinogen-like domain since fibronectin has affinity for fibrinogen (28).

The hexabrachions demonstrate how skilled the genome is in getting the most out of each piece of information. They show us exon shuffling and exon reduplication as well as divergence of reduplicated exons in the evolution of a gene and alternate splicing in generation of diversity among proteins derived from this gene.

1. Gulcher, J. R., Marton, L. S. & Stefansson, K. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2118–2122.
2. Kruse, J., Keilhauer, G., Faissner, A., Timpl, R. & Schachner, M. (1985) *Nature (London)* **316**, 146–148.
3. Erickson, H. P. & Inglesias, J. L. (1984) *Nature (London)* **311**, 267–269.
4. Chiquet-Ehrismann, R., Mackie, E. J., Pearson, C. A. & Sakakura, T. (1986) *Cell* **47**, 131–139.
5. Chiquet, M. & Fambrough, D. M. (1984) *J. Cell Biol.* **98**, 1937–1946.
6. Bourdon, M. A., Wikstrand, C. J., Furthmayr, H., Matthews, T. J. & Bigner, D. D. (1983) *Cancer Res.* **43**, 2796–2805.
7. Grumet, M., Hoffman, S., Crossin, K. L. & Edelman, G. M. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 8075–8079.
8. Erickson, H. P. & Taylor, H. C. (1987) *J. Cell Biol.* **105**, 1387–1394.
9. Hoffman, S., Crossin, K. L. & Edelman, G. M. (1988) *J. Cell Biol.* **106**, 519–532.
10. Hoffman, S. & Edelman, G. M. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 2523–2527.
11. Jones, F. S., Burgoon, M. P., Hoffman, S., Crossin, K. L., Cunningham, B. A. & Edelman, G. M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2186–2190.
12. Pearson, C. A., Pearson, D., Shibahara, S., Hofsteenge, J. & Chiquet-Ehrismann, R. (1988) *EMBO J.* **10**, 2977–2982.
13. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **24**, 5294–5299.
14. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1408–1412.
15. Gubler, U. & Hoffman, B. J. (1983) *Gene* **25**, 263–269.
16. Young, R. A. & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1194–1198.
17. Henikoff, S. (1984) *Gene* **28**, 351–359.
18. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
19. Tabor, S. & Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4767–4771.
20. Scharf, S. J., Horn, G. T. & Erlich, H. A. (1986) *Science* **233**, 1076–1078.
21. Devereux, J., Haeberl., P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
22. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), p. 80.
23. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
24. Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685.
25. Towbin, H., Staehelin, T. & Gordon, J. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4350–4354.
26. Pierschbacher, M. D., Hayman, E. G. & Ruoslahti, E. (1982) *Proc. Natl. Acad. Sci. USA* **80**, 1224–1227.
27. Obara, J., Kang, M. S. & Yamada, K. M. (1988) *Cell* **53**, 649–657.
28. Ruoslahti, E. (1988) *Annu. Rev. Biochem.* **57**, 375–413.
29. Künemund, V., Jungalwala, F. B., Fischer, G., Chou, D. K. H., Keilhauer, G. & Schachner, M. (1988) *J. Cell Biol.* **106**, 213–223.
30. Ranscht, B. (1988) *J. Cell Biol.* **107**, 1561–1573.
31. Moos, M., Tacke, R., Scherer, H., Teplow, D., Fah, K. & Schachner, M. (1988) *Nature (London)* **334**, 701–703.
32. Stoscheck, C. M. & King, L. E. (1986) *J. Cell. Biochem.* **31**, 135–152.
33. Südhof, T. C., Russell, D. W., Goldstein, J. L., Brown, M. S., Sanchez-Pescador, R. & Bell, G. I. (1985) *Science* **228**, 893–895.
34. Doolittle, R. F. (1983) *Annu. Rev. Biochem.* **53**, 195–229.