Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 1905–1909, March 1989
Developmental Biology

A detailed structural model of cytotactin: Protein homologies, alternative RNA splicing, and binding regions

(cell–substrate adhesion/epidermal growth factor-like repeats/fibronectin type III repeats/fibrinogen similarities/cDNA sequencing)

FREDERICK S. JONES, STANLEY HOFFMAN, BRUCE A. CUNNINGHAM, AND GERALD M. EDELMAN

The Rockefeller University, 1230 York Avenue, New York, NY 10021

Contributed by Gerald M. Edelman, December 23, 1988

ABSTRACT A combination of cDNA sequencing of the complete coding region, protein comparisons, binding site mapping, and electron microscopic imaging has permitted the formulation of a structural model of cytotactin. Cytotactin is a large extracellular matrix glycoprotein that displays a restricted tissue distribution during development. Although there appears to be a single cytotactin gene, multiple cytotactin polypeptides and mRNAs are detected in a variety of tissues. We report here the sequences and relationships of cDNAs that encode the complete amino acid sequences of two cytotactin polypeptides in chicken brain. The translated cDNA sequences agree with those obtained by direct analysis of cytotactin and fragments of the molecule. All regions of the polypeptides appear to be identical except for a 273 amino acid segment found in the larger but not in the smaller. At their amino termini, both polypeptides contain a cysteine-rich segment that probably includes those residues that link monomers into hexamers. This segment is followed by 13 epidermal growth factor-like (EGFL) repeats and then 8 consecutive segments that each resemble the type III repeats found in fibronectin. At their carboxyl termini, the polypeptides are similar to the β and γ chains of fibrinogen, including a calcium-binding segment. The additional sequence in the large polypeptide is inserted after the fifth type III repeat and includes three additional type III repeats. On RNA transfer blot analyses, cytotactin cDNA probes detected a 6.4-kilobase (kb) component in both brain and gizzard and larger mRNAs in both tissues, but those in gizzard were larger by about 1 kb than those in brain. A probe specific to the insert did not hybridize to the 6.4-kb mRNA in either tissue but detected the larger mRNAs in both tissues. At least a portion of the insert is thus present in both tissues, but there may be additional inserts in the gizzard mRNAs. The proposed model of cytotactin specifies the orientation of the polypeptides, the localization of interchain disulfide bonds, the structural elements constituting the thin and thick segments (EGFL repeats and type III repeats, respectively), the terminal fibrinogen-like nodular region, and the relative location of the cell-binding region.

The extracellular matrix protein cytotactin is involved in both neural and nonneural cellular interactions and histogenesis (1, 2). *In vitro*, it binds to a variety of cell types and, in some cases, causes cells to round up and inhibits their migration (3–5). The molecule has a restricted distribution that changes during development (1, 3). In the chick embryo, it appears first at gastrulation and is later expressed in distinctive patterns in the basement membrane of the neural tube and in neural crest cell migration pathways (3). The interaction of cytotactin with other extracellular matrix proteins, including fibronectin and a chondroitin sulfate proteoglycan [cytotactin-binding (CTB) proteoglycan] appears to modulate the ability of these molecules to bind cells (6).

As isolated from embryonic brain, cytotactin contains a 220-kDa component and components appearing as a 190/200-kDa doublet, all of which are closely related but nonidentical polypeptides (7). A 250-kDa form containing covalently attached chondroitin sulfate is found in brain, whereas in gizzard a 240-kDa polypeptide replaces the 220-kDa polypeptide (6, 8). In electron microscopic analyses, cytotactin appears as a six-armed structure (designated a hexabrachion) with a central core (9); biochemical analyses indicate that this is a disulfide-linked oligomer (4, 10). The cell-binding site in cytotactin as well as the sites for interaction with fibronectin and CTB proteoglycan are associated with the distal portions of the arms (4, 5). Molecules known as brachionectin and tenascin, with properties very similar to those of cytotactin (5, 10), give similar electron microscopic images and are probably identical proteins.

In initial experiments (8), a cDNA clone that accounted for about half of a single cytotactin polypeptide was shown to be similar to three different proteins: the 5′ end coded for four epidermal growth factor-like (EGFL) repeats; the following region encoded eight consecutive segments that resembled the type III repeats in fibronectin; and the 3′ end encoded a segment similar to the β and γ chains of fibrinogen. Hybridization with specific cDNA probes revealed multiple mRNAs for cytotactin but only a single gene. Subsequent partial sequence analysis of tenascin cDNAs from chick fibroblasts showed that the sequence of tenascin is identical to that of cytotactin and revealed nine additional units in the array of EGFL repeats (11).

We report here the sequences of additional cDNA clones that encode two complete polypeptides found in chicken brain and reveal an additional segment of 273 amino acids that is inserted into larger polypeptides by alternative RNA splicing.* Coupled with previous structural and binding data, these results have allowed us to formulate a detailed model of the cytotactin molecule.

MATERIALS AND METHODS

cDNA clones were isolated from three independent libraries prepared with embryonic day 10 chicken poly(A)$^+$ RNA: (*i*) a whole embryo library in λgt11 from Clontech, (*ii*) a brain library in λgt10 kindly provided by Hidesaboro Hanafusa of The Rockefeller University, and (*iii*) a brain library in λgt11 prepared in this laboratory. Libraries were screened (12) by using ^{32}P-labeled inserts from cytotactin cDNA clones pEC802 and pEC803 (8), or 50- to 60-mer synthetic oligonucleotide probes.

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

Abbreviations: EGFL, epidermal growth factor-like; CTB proteoglycan, cytotactin-binding proteoglycan.

*The sequence reported in this paper is being deposited in the EMBL/GenBank data base (accession no. J04519).

```
                                      GAATTCCGGCTGATCTGACCAGTGTGCCGCACTGTCA
AACCCTCCTTTCACACACGCGCGCACCAAATGAGACGGCACAACTTCTCTGAGTTTTGACAGG
ACGGCGAGGAATCCGGGAGCCGACAGCTGCTGCTGCAGTACCTCTGCTTCGTGGAGGCTGCCC
GTGGCAGGATCTGATCCGTCAGCCCACACGAGAATAAGCGTGCCAAGAAAGGAAAGGAAACTC
AACTTAGTTTGAACTGGCTCTCAAATTTCTCCTCCAGTCTACAAAGGCCAAACAAATATAAGA
                                     MetGlyLeuProSerGlnValLeuAlaCysAlaIleLeu
CTCCATCAGCTTTGAAGCACTACAATGGGACTCCCTTCCCAGGTTTTGGCCTGTGCCATCTTA
                                                             ▼1
GlyLeuLeuTyrGlnHisAlaSerGlyGlyLeuIleLysArgIleIleArgGlnLysArgGlu
GGTTTGCTGTACCAGCATGCCAGTGGTGGGCTCATCAAGCGAATTATCCGGCAGAAGAGGGAG
             ●
ThrGlyLeuAsnValThrLeuProGluAspAsnGlnProValValPheAsnHisValTyrAsn
ACTGGGCTCAATGTGACCTTACCAGAGGATAATCAGCCTGTGGTTTTCAATCATGTCTACAAC

IleLysLeuProValGlySerLeuCysSerValAspLeuAspThrAlaSerGlyAspAlaAsp
ATCAAGCTGCCTGTTGGCTCCCTTTGCTCTGTGGACCTGGACACAGCAAGCGGGGACGCAGAC

LeuLysAlaGluIleGluProValLysAsnTyrGluGluHisThrValAsnGluGlyAsnGln
CTGAAGGCAGAAATTGAGCCTGTCAAGAATTACGAGGAGCATACGGTGAATGAGGGGAACCAG

IleValPheThrHisArgIleAsnIleProArgArgAlaCysGlyCysAlaAlaAlaProAsp
ATTGTCTTCACGCACCGCATCAACATTCCCCGCCGGGCCTGTGGCTGTGCGGCTGCCCCAGAC
                                        100
IleLysAspLeuLeuSerArgLeuGluGluLeuGluGlyLeuValSerSerLeuArgGluGln
ATCAAGGACCTGCTGAGCAGACTGGAGGAGCTGGAGGGGCTGGTATCCTCCCTCCGGGAGCAG
      ○
CysAlaSerGlyAlaGlyCysCysProAsnSerGlnThrAlaGluGlyArgLeuAspThrAla
TGTGCCAGCGGGGCTGGATGCTGTCCTAATTCCCAGACAGCAGAAGGTCGCCTGGACACGGCC
         ○                                     ►EGFL
ProTyrCysSerGlyHisGlyAsnTyrSerThrGluIleCysGlyCysValCysGluProGly
CCCTATTGCAGTGGGCACGGCAACTACAGCACCGAGATCTGTGGCTGCGTGTGCGAGCCAGGC
                   ►1
TrpLysGlyProAsnCysSerGluProAlaCysProArgAsnCysLeuAsnArgGlyLeuCys
TGGAAAGGCCCCAACTGCTCCGAACCGGCCTGCCCACGCAACTGCCTCAACCGCGGCCTCTGC
                                             ►2
ValArgAlaLysCysIleCysGluGluGlyPheThrGlyGluAspCysSerGlnAlaArgCys
GTGCGGGCAAAGTGCATCTGCGAGGAGGGCTTTACCGGCGAGGACTGCAGCCAGGCTCGCTGC
                            200
ProSerAspCysAsnAspGlnGlyLysCysValAspGlyValCysValCysPheGluGlyTyr
CCGTCTGACTGCAACGACCAAGGCAAGTGTGTGGATGGGGTGTGCGTCTGCTTCGAGGGCTAC
                ►3
ThrGlyProAspCysGlyGluGluLeuCysProHisGlyCysGlyIleHisGlyArgCysVal
ACGGGCCCGGACTGCGGCGAGGAGCTCTGCCCCCACGGGTGTGGCATTCACGGGCGCTGTGTG
                                             ►4
GlyGlyArgCysValCysHisGluGlyPheThrGlyGluAspCysAsnGluProLeuCysPro
GGTGGACGCTGTGTGTGCCACGAGGGCTTCACTGGCGAGGACTGCAATGAGCCCCTGTGCCCC

AsnAsnCysHisAsnArgGlyArgCysValAspAsnGluCysValCysAspGluGlyTyrThr
AACAACTGTCACAACCGCGGGCGCTGTGTGGACAACGAGTGCGTCTGCGATGAGGGCTACACG
             ►5
GlyGluAspCysGlyGluLeuIleCysProAsnAspCysPheAspArgGlyArgCysIleAsn
GGAGAGGACTGCGGCGAGCTGATTTGCCCCAATGACTGCTTTGACCGCGGGCGCTGCATCAAT
            300                          ►6
GlyThrCysPheCysGluGluGlyTyrThrGlyGluAspCysGlyGluLeuThrCysProAsn
GGGACCTGCTTCTGCGAGGAGGGCTACACTGGAGAGGACTGCGGCGAGCTGACCTGCCCCAAC

AsnCysAsnGlyAsnGlyArgCysGluAsnGlyLeuCysValCysHisGluGlyPheValGly
AACTGCAACGGCAACGGGCGCTGTGAGAACGGGCTGTGTGTGTGCCATGAGGGCTTCGTGGGG
          ►7
AspAspCysSerGlnLysArgCysProLysThrCysAsnAsnArgGlyArgCysValAspGly
GATGACTGCAGCCAGAAGAGGTGCCCGAAGACGTGCAATAACCGCGGGCGCTGCGTGGATGGG
                                       ►8
ArgCysValCysHisGluGlyTyrLeuGlyGluAspCysGlyGluLeuArgCysProAsnAsp
CGCTGTGTGTGCCATGAGGGGTACCTGGGGGAGGACTGTGGGGAGCTGCGGTGCCCCAACGAC
                                                               400
CysHisAsnArgGlyArgCysIleAsnGlyGlnCysValCysAspGluGlyPheIleGlyGlu
TGCCACAACCGCGGGCGCTGCATCAACGGGCAGTGTGTGTGTGATGAGGGATTCATTGGGGAG
      ►9
AspCysGlyGluLeuArgCysProAsnAspCysGlnGlnArgGlyArgCysIleAsnGlyGln
GACTGTGGAGAGCTGCGGTGCCCCAACGACTGCCAGCAACGCGGGCGCTGCATCAATGGGCAG
                                    ►10
CysGluCysHisGluGlyPheIleGlyGluAspCysGlyGluLeuArgCysProAsnAspCys
TGCGAGTGCCACGAGGGATTCATCGGGGAGGACTGCGGGGAGCTGCGGTGTCCCAACGACTGC

AsnSerHisGlyArgCysValAsnGlyGlnCysValCysAspGluGlyTyrThrGlyGluAsp
AACAGCCATGGGCGCTGCGTCAATGGGCAGTGCGTGTGTGATGAGGGGTACACAGGGGAGGAC
   ►11
CysGlyGluLeuArgCysProAsnAspCysHisAsnArgGlyArgCysValGluGlyArgCys
TGCGGGGAGTTGCGGTGCCCCAACGACTGCCACAACCGCGGGCGCTGCGTGGAGGGACGCTGT
                                 ►12           500
ValCysAspAsnGlyPheMetGlyGluAspCysGlyGluLeuSerCysProAsnAspCysHis
GTGTGTGACAACGGCTTCATGGGGGAGGACTGCGGGGAGCTGTCCTGTCCCAATGACTGCCAC

GlnHisGlyArgCysValAspGlyArgCysValCysHisGluGlyPheThrGlyGluAspCys
CAGCACGGGCGCTGCGTCGATGGGCGCTGCGTGTGCCACGAGGGCTTCACTGGGGAAGACTGC
►13
ArgGluArgSerCysProAsnAspCysAsnAsnValGlyArgCysValGluGlyArgCysVal
CGGGAACGGTCCTGCCCCAATGACTGCAACAACGTGGGCCGCTGTGTCGAGGGACGGTGTGTC
                         EGFL◄   ►FN-I
CysGluGluGlyTyrMetGlyIleAspCysSerAspValSerProProThrGlyLeuThrVal
TGTGAGGAAGGTTACATGGGGATCGACTGTTCTGATGTGTCTCCTCCAACGGGACTGACTGTA
    ●
ThrAsnValThrAspLysThrValAsnLeuGluTrpLysHisGluAsnLeuValAsnGluTyr
ACGAATGTAACAGATAAAACGGTAAATCTGGAATGGAAGCATGAGAATCTCGTCAATGAGTAC
                              600                                ●
LeuValThrTyrValProThrSerSerGlyGlyLeuAspLeuGlnPheThrValProGlyAsn
CTTGTCACCTATGTCCCTACCAGCAGTGGTGGCTTAGATCTACAGTTCACCGTACCAGGAAAC

GlnThrSerAlaThrIleHisGluLeuGluProGlyValGluTyrPheIleArgValPheAla
CAGACATCTGCCACTATTCATGAGCTGGAGCCTGGTGTGGAATACTTCATCCGTGTCTTTGCA
                                                   ►II
IleLeuLysAsnLysLysSerIleProValSerAlaArgValAlaThrTyrLeuProAlaPro
ATCCTTAAAAACAAGAAAAGTATTCCAGTCAGTGCCAGAGTAGCGACATATTTGCCTGCTCCA

GluGlyLeuLysPheLysSerValArgGluThrSerValGlnValGluTrpAspProLeuSer
GAAGGTCTGAAATTCAAATCTGTTAGAGAAACGTCTGTCCAGGTGGAATGGGATCCTCTGAGC

IleSerPheAspGlyTrpGluLeuValPheArgAsnMetGlnLysLysAspAspAsnGlyAsp
ATTTCCTTTGATGGCTGGGAGCTGGTCTTTCGTAATATGCAGAAAAAGGATGATAATGGAGAC
                700
IleThrSerSerLeuLysArgProGluThrSerTyrMetGlnProGlyLeuAlaProGlyGln
ATAACCAGCAGCTTGAAAAGGCCGGAGACATCATATATGCAGCCAGGATTGGCACCAGGACAA
       ●                    ●
GlnTyrAsnValSerLeuHisIleValLysAsnAsnThrArgGlyProGlyLeuSerArgVal
CAGTATAATGTATCCCTTCATATAGTGAAAAACAATACCAGAGGACCAGGGCTATCCCGAGTG
         ►III
IleThrThrLysLeuAspAlaProSerGlnIleGluAlaLysAspValThrAspThrThrAla
ATAACCACAAAACTCGATGCCCCTAGCCAGATTGAGGCGAAAGATGTCACAGACACCACAGCT

LeuIleThrTrpSerLysProLeuAlaGluIleGluGlyIleGluLeuThrTyrGlyProLys
CTGATCACATGGTCCAAACCCTTGGCTGAAATTGAAGGCATAGAGCTCACATATGGCCCCAAG

AspValProGlyAspArgThrThrIleAspLeuSerGluAspGluAsnGlnTyrSerIleGly
GATGTTCCAGGGGACAGGACTACCATTGACCTCTCTGAGGATGAAAACCAATATTCTATTGGA
800                                                     *  *  *
AsnLeuArgProHisThrGluTyrGluTyrGluValThrLeuIleSerArgArgGlyAspMet
AACCTGAGGCCACACACAGAATATGAATATGAAGTGACACTCATTTCTCGGCGAGGGGACATG
                                 ►IV
GluSerAspProAlaLysGluValPheValThrAspLeuAspAlaProArgAsnLeuLysArg
GAGAGTGACCCTGCAAAAGAAGTCTTTGTCACAGACTTGGATGCTCCACGAAACCTGAAGCGA

ValSerGlnThrAspAsnSerIleThrLeuGluTrpLysPheSerHisAlaAsnIleAspAsn
GTGTCACAGACAGACAACAGCATTACTTTGGAGTGGAAGTTCAGCCATGCAAATATTGATAAT
```

```
TyrArgIleLysPheAlaProIleSerGlyGlyAspHisThrGluLeuThrValProLysGly
TACCGAATTAAGTTTGCTCCCATTTCTGGTGGAGACCACACTGAGCTGACAGTGCCAAAGGGC
                                                 900
AsnGlnAlaThrThrArgAlaThrLeuThrGlyLeuArgProGlyThrGluTyrGlyIleGly
AACCAAGCAACAACCAGAGCTACACTCACAGGTTTGAGACCTGGAACTGAATATGGCATTGGA
                                                          ►V
ValThrAlaValArgGlnAspArgGluSerAlaProAlaThrIleAsnAlaGlyThrAspLeu
GTGACAGCAGTGAGACAGGACAGGGAAAGTGCTCCTGCTACCATTAATGCTGGCACTGATCTT

AspAsnProLysAspLeuGluValSerAspProThrGluThrThrLeuSerLeuArgTrpArg
GATAACCCCAAGGACTTGGAAGTCAGTGACCCCACTGAAACCACCCTGTCCCTTCGCTGGAGA

ArgProValAlaLysPheAspArgTyrArgLeuThrTyrValSerProSerGlyLysLysAsn
AGACCAGTGGCCAAATTTGATCGTTACCGCCTCACTTACGTTAGCCCCTCTGGAAAGAAGAAC

GluMetGluIleProValAspSerThrSerPheIleLeuArgGlyLeuAspAlaGlyThrGlu
GAAATGGAGATCCCTGTGGACAGCACCTCTTTTATCCTGAGAGGATTAGACGCAGGGACGGAG
                                    1000
TyrThrIleSerLeuValAlaGluLysGlyArgHisLysSerLysProThrThrIleLysGly
TACACCATCAGTCTAGTGGCAGAGAAAGGCAGACACAAAAGCAAACCCACAACCATCAAGGGT
      ►Va                        ●
SerThrGluGluGluProGluLeuGlyAsnLeuSerValSerGluThrGlyTrpAspGlyPhe
TCGACTGAGGAAGAACCTGAGCTTGGAAACTTATCAGTGTCAGAGACTGGCTGGGATGGTTTC

GlnLeuThrTrpThrAlaAlaAspGlyAlaTyrGluAsnPheValIleGlnValGlnGlnSer
CAGCTCACCTGGACAGCAGCCGACGGGGCCTATGAGAACTTTGTCATTCAGGTGCAGCAGTCT
                      ●                                ●
AspAsnProGluGluThrTrpAsnIleThrValProGlyGlyGlnHisSerValAsnValThr
GACAATCCAGAAGAAACCTGGAACATTACAGTCCCCGGCGGACAGCACTCTGTGAACGTTACA
                        ●
GlyLeuLysAlaAsnThrProTyrAsnValThrLeuTyrGlyValIleArgGlyTyrArgThr
GGCCTCAAGGCCAACACACCTTATAACGTCACACTTTACGGTGTGATTCGAGGCTACAGAACC
                1100        ►Vb
LysProLeuTyrValGluThrThrThrGlyAlaHisProGluValGlyGluLeuThrValSer
AAACCCCTTTATGTTGAAACCACGACAGGAGCACACCCCGAAGTTGGTGAGCTAACCGTTTCC
                      ●
AspIleThrProGluSerPheAsnLeuSerTrpThrThrThrAsnGlyAspPheAspAlaPhe
GACATTACTCCTGAAAGCTTCAACCTTTCTTGGACGACCACCAACGGGGACTTTGACGCCTTT
                                                   ●
ThrIleGluIleIleAspSerAsnArgLeuLeuGluProMetGluPheAsnIleSerGlyAsn
ACTATTGAAATTATTGATTCTAACAGGTTGCTGGAGCCCATGGAGTTCAACATCTCAGGCAAT

SerArgThrAlaHisIleSerGlyLeuSerProSerThrAspPheIleValTyrLeuTyrGly
TCAAGAACAGCTCATATCTCAGGGCTTTCCCCCAGCACTGATTTTATTGTCTACCTCTATGGG
                                                   ►Vc
IleSerHisGlyPheArgThrGlnAlaIleSerAlaAlaAlaThrThrGluAlaGluProGlu
ATCTCTCATGGTTTCCGCACACAGGCAATAAGTGCTGCGGCTACAACAGAGGCAGAACCCGAG
 1200
ValAspAsnLeuLeuValSerAspAlaThrProAspGlyPheArgLeuThrTrpThrAlaAsp
GTGGACAACCTTCTGGTTTCAGATGCTACCCCAGACGGCTTCCGTCTGACCTGGACTGCAGAT

AspGlyValPheAspSerPheValLeuLysIleArgAspThrLysArgLysSerAspProLeu
GATGGGGTTTTCGACAGTTTTGTTCTAAAAATCAGGGATACCAAAAGGAAATCTGATCCACTG

GluLeuIleValProGlyHisGluArgThrHisAspIleThrGlyLeuLysGluGlyThrGlu
GAACTCATTGTACCAGGCCATGAGCGCACCCATGATATAACAGGGCTGAAAGAGGGCACTGAG

TyrGluIleGluLeuTyrGlyValSerSerGlyArgArgSerGlnProIleAsnSerValAla
TATGAAATTGAGCTCTATGGAGTTAGCAGTGGACGGCGCTCCCAACCCATAAATTCAGTAGCA
      ►VI                                          1300
ThrThrValValGlySerProLysGlyIleSerPheSerAspIleThrGluAsnSerAlaArg
ACCACAGTTGTGGGATCTCCCAAGGGAATCTCTTTCTCGGACATCACAGAAAACTCTGCTAGA

ValSerTrpThrProProArgSerArgValAspSerTyrArgValSerTyrValProIleThr
GTCAGCTGGACACCCCCCCGCAGCCGTGTGGATAGCTACAGGGTCTCCTATGTCCCCATCACA

GlyGlyThrProAsnValValThrValAspGlySerLysThrArgThrLysLeuValLysLeu
GGCGGCACTCCCAATGTTGTTACAGTTGATGGAAGCAAGACAAGGACAAAGCTGGTGAAGTTA

ValProGlyValAspTyrAsnValAsnIleIleSerValLysGlyPheGluGluSerGluPro
GTCCCAGGTGTAGACTACAACGTTAATATCATCTCTGTGAAAGGCTTTGAAGAAAGCGAACCC
                     ►VII
IleSerGlyIleLeuLysThrAlaLeuAspSerProSerGlyLeuValValMetAsnIleThr
ATTTCTGGAATTCTGAAAACAGCTCTGGACAGCCCGTCAGGACTGGTAGTGATGAACATTACA
                                            1400
AspSerGluAlaLeuAlaThrTrpGlnProAlaIleAlaAlaValAspAsnTyrIleValSer
GACTCGGAGGCTCTGGCAACCTGGCAGCCTGCAATTGCAGCTGTGGATAATTACATTGTCTCC

TyrSerSerGluAspGluProGluValThrGlnMetValSerGlyAsnThrValGluTyrAsp
TACTCTTCTGAGGATGAGCCAGAAGTTACACAGATGGTATCAGGAAACACAGTGGAGTACGAC

LeuAsnGlyLeuArgProAlaThrGluTyrThrLeuArgValHisAlaValLysAspAlaGln
CTGAATGGCCTTCGACCTGCGACAGAGTACACCCTGAGGGTGCATGCAGTGAAGGATGCGCAG
                                  ►VIII
LysSerGluThrLeuSerThrGlnPheThrThrGlyLeuAspAlaProLysAspLeuSerAla
AAGAGCGAGACCCTCTCCACCCAGTTCACTACAGGACTCGATGCTCCAAAAGATTTAAGTGCT

ThrGluValGlnSerGluThrAlaValIleThrTrpArgProProArgAlaProValThrAsp
ACCGAGGTTCAGTCAGAAACAGCTGTGATAACGTGGAGGCCTCCACGTGCTCCTGTCACTGAT
                        1500
TyrLeuLeuThrTyrGluSerIleAspGlyArgValLysGluValIleLeuAspProGluThr
TACCTCCTGACCTACGAGTCCATTGATGGCAGAGTCAAGGAAGTCATCCTAGACCCTGAGACG

ThrSerTyrThrLeuThrGluLeuSerProSerThrGlnTyrThrValLysLeuGlnAlaLeu
ACCTCCTACACCCTGACAGAGCTGAGCCCATCCACTCAATACACAGTGAAACTTCAGGCACTG
                                             FN◄
SerArgSerMetArgSerLysMetIleGlnThrValPheThrThrThrGlyLeuLeuTyrPro
AGCAGATCTATGAGGAGCAAAATGATCCAGACTGTTTTCACCACAACTGGTCTTCTTTATCCT
     ►FG
TyrProLysAspCysSerGlnAlaLeuLeuAsnGlyGluValThrSerGlyLeuTyrThrIle
TATCCTAAAGACTGCTCCCAAGCTCTCCTGAATGGAGAGGTCACCTCTGGGCTCTACACTATT

TyrLeuAsnGlyAspArgThrGlnProLeuGlnValPheCysAspMetAlaGluAspGlyGly
TATCTGAATGGAGACAGGACACAGCCTCTGCAAGTCTTCTGTGACATGGCTGAAGATGGAGGC
     1600
GlyTrpIleValPheLeuArgArgGlnAsnGlyLysGluAspPheTyrArgAsnTrpLysAsn
GGATGGATTGTGTTCCTGAGGCGTCAAAATGGAAAGGAAGATTTCTACAGGAACTGGAAGAAT

TyrValAlaGlyPheGlyAspProLysAspGluPheTrpIleGlyLeuGluAsnLeuHisLys
TACGTGGCCGGCTTTGGAGATCCCAAGGATGAATTCTGGATAGGTCTGGAGAACCTCCACAAA

IleSerSerGlnGlyGlnTyrGluLeuArgValAspLeuArgAspArgGlyGluThrAlaTyr
ATCAGCTCTCAGGGGCAGTACGAGCTGCGTGTGGATCTGAGAGACAGAGGTGAGACAGCCTAT

AlaValTyrAspLysPheSerValGlyAspAlaLysThrArgTyrArgLeuArgValAspGly
GCTGTGTACGACAAGTTCAGCGTTGGAGATGCCAAGACCCGGTACCGGCTGAGGGTGGATGGC
                                                            1700
TyrSerGlyThrAlaGlyAspSerMetThrTyrHisAsnGlyArgSerPheSerThrPheAsp
TACAGTGGCACAGCAGGTGACTCCATGACCTACCATAATGGAAGATCCTTCTCCACTTTTGAC

LysAspAsnAspSerAlaIleThrAsnCysAlaLeuSerTyrLysGlyAlaPheTrpTyrLys
AAGGACAATGATTCTGCTATCACCAACTGTGCTTTGTCATACAAGGGTGCTTTCTGGTACAAG

AsnCysHisArgValAsnLeuMetGlyArgTyrGlyAspAsnAsnHisSerGlnGlyValAsn
AATTGTCACCGAGTCAATCTGATGGGCAGATATGGTGACAACAACCACAGTCAGGGTGTTAAT
                                                          FG◄
TrpPheHisTrpLysGlyHisGluTyrSerIleGlnPheAlaGluMetLysLeuArgProSer
TGGTTCCACTGGAAGGGCCACGAATACTCCATCCAGTTTGCAGAGATGAAACTGAGACCCTCC

SerPheArgAsnLeuGluGlyArgArgLysArgAla 1777
AGCTTTCGGAATCTGGAAGGAAGACGAAAGCGAGCATAAAGCCCTTGGGATGGTGAAAGGGCT
ACGGGCAGGGCAACATGGGGAGGGACAGAGAGCGGGGGGCATGGAGGATCTCTGGCATCACT
GGGGTTATGGGTGTGAGGAGCTGGTAGTCGTACCAAAGCATCGCAACCCTTGGCACAAGAGCC
CAAACAACGAGCCTTACGTGTCCCAGCAATTCCAGCAGAGCAGCTCCAGCTCTGCCCACTGCT
GATGTCCTTCACGCCAAAGACAACGATCTCAAGGGTTGTATGCTGTTTTCTTCATTTTTCTTT
TCTCAGCCTCTGGGATGAAGTTCTTCCACGGCGGAATTC
```

FIG. 1. *(Legend appears at the bottom of the opposite page.)*

Recombinant bacteriophage DNA inserts were subcloned as *Eco*RI fragments into Bluescript vectors (8), or as small fragments generated by digestion with *Hpa* II, *Sau*3AI, *Taq* I, and *Hae* III and ligated into compatibly cut M13mp18 and M13mp19 vectors (13). Subclones were sequenced (14) and the data were compiled by using the Staden ANALSEQ programs (15) and used in data base[†] searches (16).

Protein sequence analysis by automated Edman degradation was performed on intact cytotactin, chymotryptic fragments, and specific CNBr fragments electroeluted from gels. Protein sequence determinations and oligonucleotide syntheses were conducted in The Rockefeller University Sequencing Facility.

RESULTS

The previously described cytotactin cDNA clone, λC801, contained two *Eco*RI fragments, 2.0 kilobases (kb) (pEC802) and 0.8 kb (pEC803) long, and encoded 933 amino acids, or about half of a cytotactin polypeptide (8). An additional clone, subsequently isolated by antibody screening, contained two *Eco*RI fragments, one 3.8 kb long and another identical to pEC803. The 3.8-kb fragment contained all of pEC802 but extended further 5′ and included an additional insert of 819 base pairs, suggesting that there are at least two forms of cytotactin in embryonic chick brain, one having the insert and one lacking it. The new clone was sequenced to confirm and amend sequences of pEC802 and -803. A synthetic oligonucleotide corresponding to the most 5′ portion of the 3.8-kb clone was used to select clones at the 5′ end of cytotactin, and pEC803 was used to select clones that completed the sequence at the 3′ end.

The cDNA and deduced amino acid sequence of cytotactin are shown in Fig. 1. Five single base pair changes amended previously reported sequences (8) at amino acids 530–538, 988–1011, 1285–1289, and 1371–1380. The sequence differs at six positions from that reported for tenascin (11): amino acids 171, 189, 348, 412–413, 417, and 690. Of the 6061 sequenced base pairs, 5430 code for protein, with 313 base pairs of 5′ untranslated and 318 base pairs of 3′ untranslated mRNA sequence. The larger form of cytotactin contains 1777 amino acids (M_r = 198,313) and the smaller 1504 (M_r = 167,846). There are 33 additional amino acids in the precursor sequence; the first 23 correspond to a typical signal peptide (17). The remaining 10 include basic residues similar to those found in the precursor segments of calcium-dependent cell adhesion molecules, such as L-CAM (18).

The reading frame was verified by protein sequence analysis of intact cytotactin and fragments derived from it. The amino-terminal sequence of the mature peptide was determined directly to residue 21. A chymotryptic fragment of cytotactin from the disulfide bonded region of the molecule [fraction I(4)] begins at amino acid 68; a phenylalanine is at position 67, consistent with the specificity of the enzyme. The mixture of the 90-kDa and 65-kDa chymotryptic fragments of cytotactin that directly mediates cell attachment [fraction II(4)] gave a single sequence, Xaa-Xaa-Xaa-Leu-Asp-Ala-Pro, which appears in amino acids 738 to 744, 830 to 836, and 1460 to 1466. These fragments probably do not begin at residue 1460 because the remainder of the molecule (318 amino acids) is not large enough to account for either the 90-kDa or the 65-kDa fragment.

A 75-kDa CNBr fragment that also contains interchain disulfide bonds (4) begins at position 121. The sequence is not preceded by a methionine, however, so it is presently unclear how this fragment (which was obtained in high yield) was generated. A 35-kDa CNBr fragment of cytotactin (4) is located within the additional insert beginning at amino acid 1150; as expected, residue 1149 is methionine. Sequence analysis of a mixture of two 17-kDa CNBr fragments obtained by $NaDodSO_4$/polyacrylamide gel electrophoresis gave a major sequence beginning at residue 1593 and a minor sequence beginning at residue 1421; in each case the previous residue is methionine.

RNA transfer blots (Fig. 2) using the 3.8-kb *Eco*RI fragment or the insert from pEC803 as probes revealed at least two messages (6.4 and 7.2 kb) in embryonic day 15 brain, and at least three messages (6.4, 8.0, and 8.3 kb) in gizzard. However, a 293-base-pair *Pst* I–*Hin*dIII fragment derived from the additional insert detected only the 7.2-kb mRNA in brain and only the 8.0- and 8.3-kb mRNAs in gizzard. In both tissues the 6.4-kb message was not detected by this probe.

DISCUSSION

A detailed model of cytotactin in terms of the similarities to other proteins, the orientation and interaction of the polypeptides, and the elements that correspond to features seen in electron micrographs is presented in Fig. 3. The amino-terminal portion contains the cysteines linking the six polypeptides to each other. Next are 13 EGFL repeats that probably make up the thinner portions (9) of the hexabrachion arms, followed by 8 or 11 segments that resemble fibronectin type III repeats and that may comprise the thicker portion of each arm (see Fig. 3*B*). The terminal fibrinogen-like segment probably corresponds to the nodular region at the distal end of each arm.

The amino-terminal 142 amino acids do not resemble any other known protein. This segment includes eight cysteine residues, which are sufficient to link both cytotactin monomers into trimers (10) and trimers into hexamers. The linkage is assumed to be direct, but it is possible that other small molecules serve as linkers in either trimer or hexamer formation.

In the carboxyl-terminal direction, the next segment of 415 amino acids includes 13 complete EGFL repeats of 31 amino acids each, preceded by 12 amino acids that comprise an incomplete EGFL repeat. The EGFL repeats are very similar to each other: 10 of the 31 amino acids are identical in every repeat and 80% of the residues match a consensus sequence. A variety of proteins contain EGFL repeats. The repeats in cytotactin most resemble those found in the Notch gene product of *Drosophila melanogaster* (19) and the B1 chain of laminin (20), but they also resemble cysteine-rich regions in endothelial glycoprotein IIIa (21) and the β subunit of integrin (22). By analogy to the known structure of human EGF (23) all of the cysteines in each cytotactin repeat are probably involved in intrachain bonds. Relative to the EGFL domains in other

[†]EMBL/GenBank Genetic Sequence Database (1988) GenBank (Bolt, Beranek, and Newman Laboratories, Cambridge, MA), Tape Release 56; and Protein Identification Resource (1988) Protein Sequence Database (Natl. Biomed. Res. Found., Washington, DC), Release 17.

FIG. 1 *(on opposite page)*. DNA and deduced amino acid sequence of cytotactin. Residues identified by protein sequencing are underlined. The amino terminus of the mature protein is marked by ▼ and *Eco*RI linkers used to construct the libraries are boxed. Potential asparagine-glycosylation sites (●), cysteines not in EGFL repeats (boxes), the Arg-Gly-Asp sequence (* * *), and potential sites for addition of glycosaminoglycans (○) are indicated. EGFL repeats, fibronectin type III repeats (FN), and the fibrinogen homology region (FG) are demarcated by arrows; specific repeats are designated with arabic numerals (EGFL) or roman numerals (type III). The additional domain found in the VaVbVc form of cytotactin is enclosed in a large box. The amino acids similar to those in the calcium-binding site of fibrinogen are underscored with a broken line. Amended DNA and corresponding amino acid sequences that were incorrect in pEC802 and -803 are in italics.

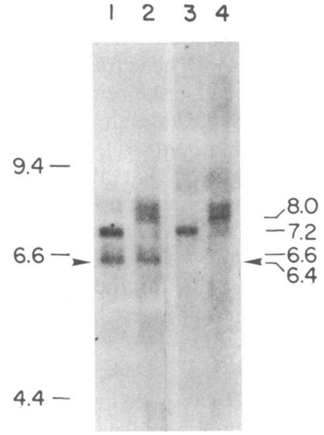

FIG. 2. Blot analysis of cytotactin-specific mRNAs. Two micrograms of poly(A)$^+$ RNA from the indicated organs was resolved in a 0.8% agarose/formaldehyde gel, transferred to nitrocellulose, hybridized to either the ^{32}P-labeled 3.8-kb *Eco*RI fragment (lanes 1 and 2) or a ^{32}P-labeled 293-base-pair *Pst* I/*Hin*dIII fragment from the VaVbVc region (lanes 3 and 4), and washed (8). Lanes 1 and 3, embryonic day 15 brain; lanes 2 and 4, embryonic day 15 gizzard. The positions and lengths (in kb) of denatured *Hin*dIII fragments of phage λ DNA are indicated. The arrow demarcates the 6.4-kb mRNA not detected by the VaVbVc probe.

proteins, those in cytotactin appear to be compact and may separate the arms of the hexabrachion by providing rigidity. In some molecules, such repeats bind calcium (24), but those in cytotactin lack the necessary aspartic acid residues.

The next region of cytotactin is composed of units similar to fibronectin type III repeats (25) which have been identified in several cell adhesion molecules (26–28). In fibronectin and cytotactin, alternative RNA splicing yields various forms of the molecules having different numbers of these units (25). The smaller cytotactin polypeptide contains eight repeats, whereas the larger contains 11 repeats. The additional type III repeats are inserted just after the fifth type III repeat (V) and are more similar to each other than to the eight repeats present in both forms of the molecule. This insert, designated VaVbVc, can account for the difference in molecular mass between the 190/200-kDa and the 220-kDa components of cytotactin in brain, but the differences between the components in the 190/200-kDa doublet remain to be determined. At least part of VaVbVc, which was found in its entirety in a brain cDNA clone, is present in the still larger gizzard mRNAs; the nature of any additional inserts in gizzard mRNA is unknown.

We have recently mapped (4) the binding functions of cytotactin to a mixture of two closely related chymotryptic fragments of cytotactin (90 and 65 kDa). From their partial sequence these fragments appear to begin at amino acid 738 or 830 within the type III repeats. Only residue 830 is preceded by an optimal residue for chymotryptic cleavage (phenylalanine), but a fragment generated at this site would lack the Arg-Gly-Asp sequence (position 817–819) whereas cleavage at residue 737 would include it. Arg-Gly-Asp-containing peptides which inhibit cell binding to fibronectin (29) also inhibit the cell binding activity of the chymotryptic fragments (4) but they might do so indirectly—e.g., by

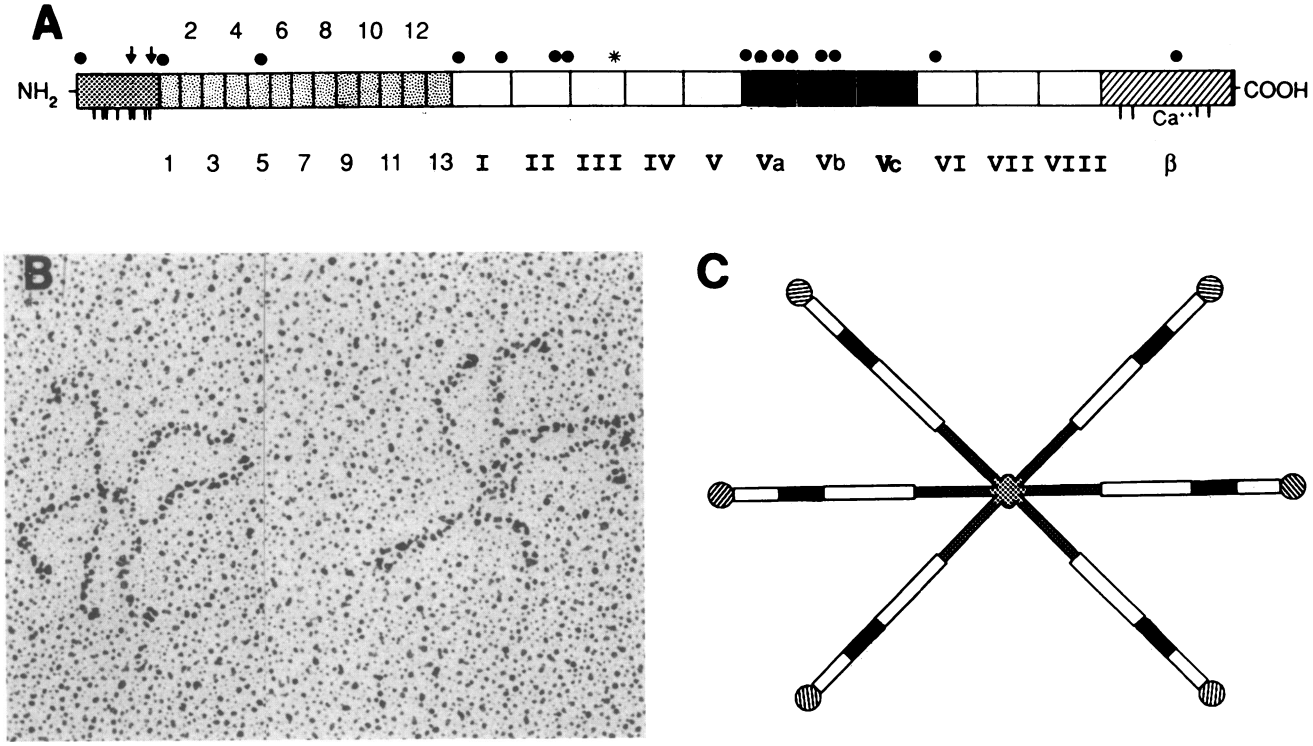

FIG. 3. Functional map of cytotactin. (*A*) Schematic drawing of the cytotactin polypeptides. The EGFL repeats (lightly stippled boxes) are numbered 1–13 and type III repeats (open boxes) are designated by roman numerals with those in the additional insert (black boxes) designated Va, Vb, and Vc. The region similar to the β subunit of fibrinogen (diagonal lines and β symbol) includes a putative calcium-binding domain (Ca^{++}). Darkly stippled boxes denote regions that have no extensive homology to any known protein, and they include the amino-terminal 142 amino acids and the carboxyl-terminal 13 amino acids. Cysteine residues that are not in EGFL repeats (|), potential asparagine-glycosylation sites (●), potential glycosaminoglycan addition sites (↓), and the Arg-Gly-Asp sequence (*) are indicated. (*B*) Electron micrograph (courtesy of Joseph W. Becker, The Rockefeller University) of chicken brain cytotactin showing two hexabrachions. (*C*) Proposed model of a cytotactin hexamer. The amino-terminal regions of each polypeptide are disulfide-linked and represented as a single structure forming the core of the hexabrachion. The EGFL repeats are cross-hatched; the type III repeats present in both polypeptides are white, whereas those in the insert are black; the fibrinogen-like nodular region is marked with slanted lines. The EGFL repeats are assumed to make up the thin parts of the arms and the type III repeats the thicker portions.

interacting with a cytotactin receptor and blocking its ability to bind to a non-Arg-Gly-Asp sequence in cytotactin (for example, see ref. 30).

Antibodies against a 35-kDa CNBr fragment recognize both the 90- and 65-kDa chymotryptic fragments and inhibit the binding of cytotactin to cells, to fibronectin, and to CTB proteoglycan (4). This CNBr fragment begins in the polypeptide encoded by the VaVbVc insert and extends beyond it, suggesting that the 90-kDa fragment may contain the insert and the 65-kDa fragment may not. This CNBr fragment does lack an Arg-Gly-Asp sequence, but the antibodies are large enough to inhibit binding to sites in other parts of the molecule that may include the Arg-Gly-Asp sequence. Alternatively, two or more sites may be involved in cytotactin's cell-binding activity, as has been shown recently (31) for the binding and spreading of fibroblasts on fibronectin.

The last domain of cytotactin contains 207 amino acids that resemble both the β and γ chains of fibrinogen and the fibrinogen-like sequence encoded by the cDNA clone pT49 (38). The similarity to fibrinogen ends at amino acid 1764, which corresponds to the carboxyl terminus of the β chain. The γ chain contains additional residues (32), but the remainder of cytotactin (13 amino acids) is not similar to this sequence. The sequence from amino acid 1560 to amino acid 1590 of cytotactin resembles the intrachain disulfide loop in fibrinogen (32), and amino acids 1699 to 1713 are similar to calcium-binding sites in fibrinogen, calmodulin, and thrombospondin (33–35). The activity of cytotactin is Ca^{2+}-dependent (6), and we have recently found that native cytotactin on nitrocellulose binds $^{45}Ca^{2+}$ (data not shown).

The cytotactin sequence predicts nine possible asparagine-glycosylation sites in the smaller polypeptide, with six additional sites in the region unique to the larger polypeptide. Consistent with the observation that cytotactin can contain covalently bound chondroitin sulfate (6), the sequence includes two sites (Ser-Gly-Xaa-Gly) for the addition of glycosaminoglycan side chains (36). These sites (amino acids 109 and 131) immediately precede the EGFL domain.

Cytotactin is composed almost entirely of units (EGFL repeats, fibronectin type III repeats, fibrinogen-like regions) that are also present in other well-characterized proteins. Although each type of unit has been found in a variety of proteins, to our knowledge cytotactin is the only protein described to date to display all three. The evolution of complex arrays of duplicated elements from apparently diverse sources appears to be a common theme in both cell–cell and cell–substrate adhesion molecules. In addition, the genes for both classes of molecules give rise to a variety of polypeptides by alternative RNA splicing. Alternative RNA splicing of the VaVbVc insert, which includes three type III repeats, produces forms of cytotactin that are expressed later in development than forms lacking this insert (6, 8). Moreover, this domain is inserted in the portion of the molecule that binds to cells, to fibronectin, and to CTB proteoglycan (4). The entire binding domain in its variant forms in a multivalent structure may lead to global modulation (37) accounting for the rounding up and relative immobilization of cells bound to the molecule (3–5). Such modulation may alter the behavior of cells in contact with cytotactin-containing matrices in a temporally and a spatially controlled manner during embryogenesis.

We thank Dr. Joseph Becker for providing the electron micrograph of chicken brain cytotactin. We are grateful to Mses. Margaret Mueller, Carin Gonsalves, Kim Drozdoski, Talia Spanier, and Ruth Arriaga for excellent technical assistance. This work was supported by U.S. Public Health Service Grants HL 37641, HD 16550, DK 04256, and HD 09635 and a Senator Jacob Javits Center of Excellence in Neuroscience Grant (NS 22789). F.S.J. is a Fellow of the Lucille P. Markey Charitable Trust.

1. Crossin, K. L., Hoffman, S., Grumet, M., Thiery, J.-P. & Edelman, G. M. (1986) *J. Cell Biol.* **102,** 1917–1930.
2. Chiquet-Ehrismann, R., Mackie, E. J., Pearson, C. A. & Sakahura, T. (1986) *Cell* **47,** 131–139.
3. Tan, S.-S., Crossin, K. L., Hoffman, S. & Edelman, G. M. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 7977–7981.
4. Friedlander, D. R., Hoffman, S. & Edelman, G. M. (1988) *J. Cell Biol.* **107,** 2329–2340.
5. Chiquet-Ehrismann, R., Kalla, P., Pearson, C. A., Beck, K. & Chiquet, M. (1988) *Cell* **53,** 383–390.
6. Hoffman, S., Crossin, K. L. & Edelman, G. M. (1988) *J. Cell Biol.* **106,** 519–532.
7. Grumet, M., Hoffman, S., Crossin, K. L. & Edelman, G. M. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 8075–8079.
8. 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.
9. Erickson, H. P. & Iglesias, J. L. (1984) *Nature (London)* **311,** 267–269.
10. Erickson, H. P. & Taylor, H. C. (1987) *J. Cell Biol.* **105,** 1387–1394.
11. Pearson, C. A., Pearson, D., Shibahara, S., Hofsteenge, J. & Chiquet-Ehrismann, R. (1988) *EMBO J.* **7,** 2977–2982.
12. Young, R. A. & Davis, R. W. (1983) *Science* **222,** 778–782.
13. Messing, J. (1983) *Methods Enzymol.* **101,** 20–78.
14. Tabor, S. & Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 4767–4771.
15. Staden, R. (1982) *Nucleic Acids Res.* **10,** 2951–2961.
16. Wilbur, W. J. & Lipman, D. J. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 726–730.
17. Von Heijne, G. (1986) *Nucleic Acids Res.* **14,** 4683–4690.
18. Sorkin, B. C., Hemperly, J. J., Edelman, G. M. & Cunningham, B. A. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 7617–7621.
19. Wharton, K. A., Johansen, K. M., Xu, T. & Artavanis-Tsakonas, S. (1985) *Cell* **43,** 567–581.
20. Sasaki, M., Kato, S., Kohno, K., Martin, G. R. & Yamada, Y. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 935–939.
21. Fitzgerald, L. A., Steiner, B., Rall, S. C., Jr., Lo, S. & Phillips, D. R. (1987) *J. Biol. Chem.* **262,** 3936–3939.
22. Hynes, R. O. (1987) *Cell* **48,** 549–554.
23. Cooke, R. M., Wilkinson, A. J., Baron, M., Pastore, A., Tappin, M. J., Campbell, I. D., Gregory, H. & Sheard, B. (1987) *Nature (London)* **327,** 339–341.
24. Fullmer, C. S. & Wasserman, R. H. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 4772–4776.
25. Kornblihtt, A. R., Vibe-Pedersen, K. & Baralle, F. E. (1984) *EMBO J.* **3,** 221–226.
26. Moos, M., Tacke, R., Scherer, H., Teplow, D., Früh, K. & Schachner, M. (1988) *Nature (London)* **334,** 701–703.
27. Streuli, M., Krueger, N. X., Hall, L. R., Schlossman, S. F. & Saito, H. (1988) *J. Exp. Med.* **168,** 1553–1562.
28. Ranscht, B. (1988) *J. Cell Biol.* **107,** 1561–1574.
29. Pierschbacher, M. D. & Ruoslahti, E. (1984) *Nature (London)* **309,** 30–33.
30. Santoro, S. A. & Lawing, W. J., Jr. (1987) *Cell* **48,** 867–873.
31. Obara, M., Kang, M. S. & Yamada, K. M. (1988) *Cell* **53,** 649–657.
32. Doolittle, R. F. (1984) *Annu. Rev. Biochem.* **53,** 195–229.
33. Dang, C. V., Ebert, R. F. & Bell, W. R. (1985) *J. Biol. Chem.* **260,** 9713–9719.
34. Putkey, J. A., Ts'ui, K. F., Tanaka, T., Lagace, L., Stein, J. P., Lai, E. C. & Means, A. R. (1983) *J. Biol. Chem.* **258,** 11864–11870.
35. Lawler, J., Weinstein, R. & Hynes, R. O. (1988) *J. Cell Biol.* **107,** 2351–2362.
36. Bourdon, M. A., Krusius, T., Campbell, S., Schwartz, N. B. & Ruoslahti, E. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 3194–3198.
37. Edelman, G. M. (1976) *Science* **192,** 218–226.
38. Koyama, T., Hall, L. R., Haser, W. G., Tonegawa, S. & Saito, H. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 1609–1613.