The EMBO Journal vol.8 no.1 pp.65-72, 1989

# Amino acid sequence of mouse nidogen, a multidomain basement membrane protein with binding activity for laminin, collagen IV and cells

**Karlheinz Mann, Rainer Deutzmann, Monique Aumailley, Rupert Timpl, Laura Raimondi[1], Yoshihiko Yamada[1], Te-cheng Pan[2], Dorothy Conway[2] and Mon-Li Chu[2]**

Max-Planck-Institut für Biochemie, D-8033 Martinsried, FRG, [1]Laboratory of Developmental Biology and Anomalies, National Institute of Dental Research, National Institutes of Health, Bethesda, MD 20892, and [2]Departments of Biochemistry and Molecular Biology and Dermatology, Jefferson Institute of Molecular Medicine, Thomas Jefferson University, Philadelphia, PA 19107, USA

Communicated by R.Timpl

**The whole amino acid sequence of nidogen was deduced from cDNA clones isolated from expression libraries and confirmed to ~50% by Edman degradation of peptides. The protein consists of some 1217 amino acid residues and a 28-residue signal peptide. The data support a previously proposed dumb-bell model of nidogen by demonstrating a large N-terminal globular domain (641 residues), five EGF-like repeats constituting the rod-like domain (248 residues) and a smaller C-terminal globule (328 residues). Two more EGF-like repeats interrupt the N-terminal and terminate the C-terminal sequences. Weak sequence homologies (25%) were detected between some regions of nidogen, the LDL receptor, thyroglobulin and the EGF precursor. Nidogen contains two consensus sequences for tyrosine sulfation and for asparagine β-hydroxylation, two *N*-linked carbohydrate acceptor sites and, within one of the EGF-like repeats an Arg-Gly-Asp sequence. The latter was shown to be functional in cell attachment to nidogen. Binding sites for laminin and collagen IV are present on the C-terminal globule but not yet precisely localized.**

*Key words:* nidogen/binding activity/laminin/collagen IV

## Introduction

Nidogen is one of the ubiquitous proteins in basement membranes, extracellular matrices in close molecular and functional contact to a large variety of cells. It is known that these contacts are mainly mediated by laminin and collagen IV, two large proteins which are highly integrated in the matrix architecture (reviewed by Kleinman *et al.*, 1985; Timpl and Dziadek, 1986; Martin and Timpl, 1987). Nidogen is often found in close association with laminin (Dziadek and Timpl, 1985; Dziadek *et al.*, 1985a; Paulsson *et al.*, 1987) but its precise biological function remains to be identified.

After isolation from the matrix of the Engelbreth-Holm-Swarm (EHS) tumor, nidogen was shown to consist of a single 148 kd polypeptide chain. This chain is folded into two globular domains of unequal size which are connected by a 17 nm long rod-like domain (Paulsson *et al.*, 1986, 1987). The protein is very susceptible to degradation by endogenous proteases with cleavages occurring at both ends of the chain as demonstrated by limited sequence analysis (Dziadek *et al.*, 1985b; Paulsson *et al.*, 1986; Mann *et al.*, 1988). A related protein, entactin, was originally identified as a sulfated 158 kd polypeptide (Carlin *et al.*, 1981) and shown to contain tyrosine sulfate (Paulsson *et al.*, 1985). Entactin also binds to laminin (Hogan *et al.*, 1982; Carlin *et al.*, 1983) and was recently characterized by a partial cDNA sequence (Durkin *et al.*, 1987). This deduced amino acid sequence contained some short segments related to nidogen but in different positions on the protein (Mann *et al.*, 1988) leaving the relationship between the two proteins still an open question.

The laminin-binding site of nidogen was recently localized to its C-terminal globular domain (Mann *et al.*, 1988). In addition, we found that the same domain binds to triple helical segments of collagen IV (M.Aumailley *et al.*, in preparation) suggesting a mediator function for nidogen in basement membranes. In order to allow a better understanding of the structure-function relationships we have now determined the whole primary structure of mouse nidogen by cDNA and peptide sequencing. The data reveal a single Arg-Gly-Asp (RGD) sequence within one of several cysteine-rich motifs with similarity to epidermal growth factor (EGF). That this putative cell-binding sequence (Ruoslahti, 1988) in fact mediates cell-binding to nidogen was shown in additional experiments.

## Results

### *Cloning of cDNA encoding nidogen*

Screening of two different λgt11 expression libraries made from mouse EHS tumor mRNA by antibodies against mouse nidogen allowed the isolation of four overlapping clones (6c, 21b, 37a, 40a) with a size of 0.5-1.5 kb. They were all located upstream to an internal *Eco*RI site present within the nucleotide sequence encoding nidogen (Figure 1). Two other clones that contained sequences downstream from this *Eco*RI site were obtained from a third EHS tumor library constructed in the λZAP vector. These two clones, N-2 and N-3, started ~1000 and 400 nt, respectively, to the 5′ end of this *Eco*RI site, both extended over this site to another *Eco*RI site 240 nt to the 3′ end, and then both terminated with a short poly(A) tail (~15 nt) ~400 nt further downstream. In addition, a 3.6 kb clone N-5 was isolated from a mouse F9 cell library. This clone ended at the first internal *Eco*RI site and extended close to the 5′ end of nidogen mRNA. A restriction map of these clones is shown in Figure 1.

Northern hybridization of the nidogen probe 6c to mouse tumor RNA demonstrated a major band of ~6 kb and a weaker reacting band of ~4 kb (Figure 2B). The same probe also hybridized to a similar 6 kb band in human mRNA under stringent conditions (Figure 2A), indicating a distinct sequence similarity of mouse and human nidogen.

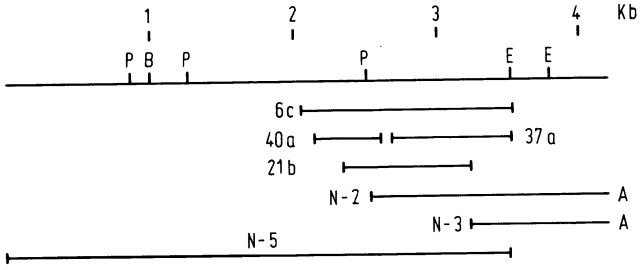


**Fig. 1.** Alignment and restriction maps of cDNA clones encoding mouse nidogen. Clones were purified from either λgt11 EHS tumor libraries (6c, 37a, 21b, 40a) after antibody screening or from an EHS tumor library in the λZAP vector (N-2, N-3) or a λgt11 F-9 cell library (N-5) after screening with 6c probe. B: *Bam*HI, E: *Eco*RI, P: *Pst*I. A denotes a poly(A) tail.

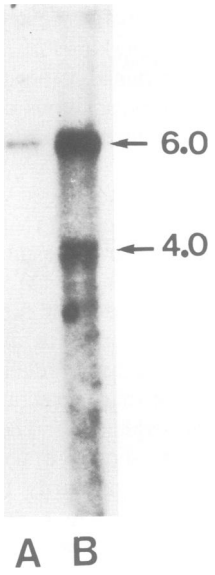


**Fig. 2.** Northern hybridization of mouse nidogen cDNA probe 6c to RNA from EHS tumor or human hepatoma cell line G2. Lanes were loaded with either total RNA from EHS tumor (20 μg, **lane B**) or with poly(A)$^+$ RNA from hepatoma (1.8 μg, **lane A**). Band sizes are indicated in kb.

### *Amino acid sequence of nidogen deduced from cDNA clones*

The various cDNA clones were used to determine most of the sequence of nidogen mRNA including the entire coding region. First, the 1.5 kb clone 6c was analyzed by the shot gun strategy in both directions and the sequence was extended to the 3′ and 5′ ends using clones N-2, N-3 and N-5 as described in Material and methods.

The analyzed cDNA sequence (4185 nt) and the deduced amino acid sequence are shown in Figure 3. Some 106 nt downstream from the 5′ end the sequence encodes the sequence LNRQELFPFG which was obtained previously at the protein level as the N-terminal sequence of nidogen (Dziadek *et al.*, 1985b; Paulsson *et al.*, 1986). This indicates that a Cys−Leu bond is the site of cleavage of the signal peptide. An open reading frame (ORF) extending 3651 nt to the first stop codon TGA at the 3′ end demonstrated that mouse nidogen consists of 1217 amino acid residues. This sequence contains 48 cysteine residues, many of them located within EGF-like repeats, a single RGD sequence possibly involved in cell-binding and two putative *N*-linked carbohydrate acceptor sequences. The nidogen sequence is preceded by a typical signal peptide (von Heijne, 1986) of either 18 or 28 amino acid residues. Ambiguity on the size of the signal peptides arises from two in-frame Met codons in nucleotide positions 22 and 52. A purine base is present 3 nt upstream from both ATG codons which is the most conserved feature for initiation of translation in vertebrates (Kozak, 1987). The sequence is terminated at the 3′ end by a 415 nt non-coding region and a poly(A) tail. This sequence contains several stop codons in all three reading frames.

### *Peptide chemistry*

About 50% of the deduced nidogen sequence was confirmed by Edman degradation of larger fragments and small tryptic peptides (Figure 3). Particular attention was given to the rod-like domain and the C-terminal globular domain because of their possible involvement in biological activities (discussed below). The rod-like domain was isolated as tryptic fragment T-40 (Mann *et al.*, 1988), cleaved after reduction and alkylation with SV8 protease and the peptides obtained were resolved by reversed-phase chromatography (Figure 4). Most of the peaks were sequenced and fitted exclusively positions 631−874 of the nidogen sequence. The C-terminal 22 kd thrombin fragment Th-22 (Mann *et al.*, 1988) was reduced and cleaved with trypsin and produced a less complex profile of peptides. All prominent peaks were sequenced and accounted for ~80% of the sequence deduced for the last 180 C-terminal amino acid residues. No peptide sequences were found beyond the predicted stop codon.

Peptide sequences for the N-terminal half of nidogen were generated in a similar way from the N-terminal thrombin fragment Th-50 (Mann *et al.*, 1988) and the endogenous fragment Nd-100 (Paulsson *et al.*, 1986) which were subjected to trypsin or elastase cleavage and HPLC separation omitting the reduction of disulfide bonds (data not shown). This allowed the generation of 14 separate peptide sequences matching N-terminal positions within nidogen. In addition, we found the N-terminal sequences determined previously (Dziadek *et al.*, 1985b; Paulsson *et al.*, 1986; Mann *et al.*, 1988) for endogenous nidogen fragments Nd-130, Nd-100, Nd-80 and Nd-40 and some thrombin fragments (Figure 3).

### *Cell-binding to nidogen mediated by an RGD sequence*

The observation of a single RGD sequence (Figure 3) prompted us to examine cell-binding properties of nidogen by cell adhesion assays (Aumailley and Timpl, 1986; Aumailley *et al.*, 1987). Studies with osteosarcoma cell line SAOS-2 showed a typical dose−response profile with a plateau region of attachment achieved at low concentrations of nidogen (Figure 5). Similar adhesion profiles were observed with human skin fibroblasts and SV-40 transformed lung fibroblasts. Cell attachment and spreading on nidogen required, however, 2−3 h when compared to the more active laminin−nidogen complex (30 min) which also caused higher attachment rates (Figure 5). The higher activity of the complex is presumably due to a high affinity binding site present on the long arm of laminin which dominates cellular interactions (Aumailley *et al.*, 1987).

A more precise localization of the cell-binding site was achieved by using nidogen fragment Nd-80 (Paulsson *et al.*, 1986) in attachment assays which showed an activity similar

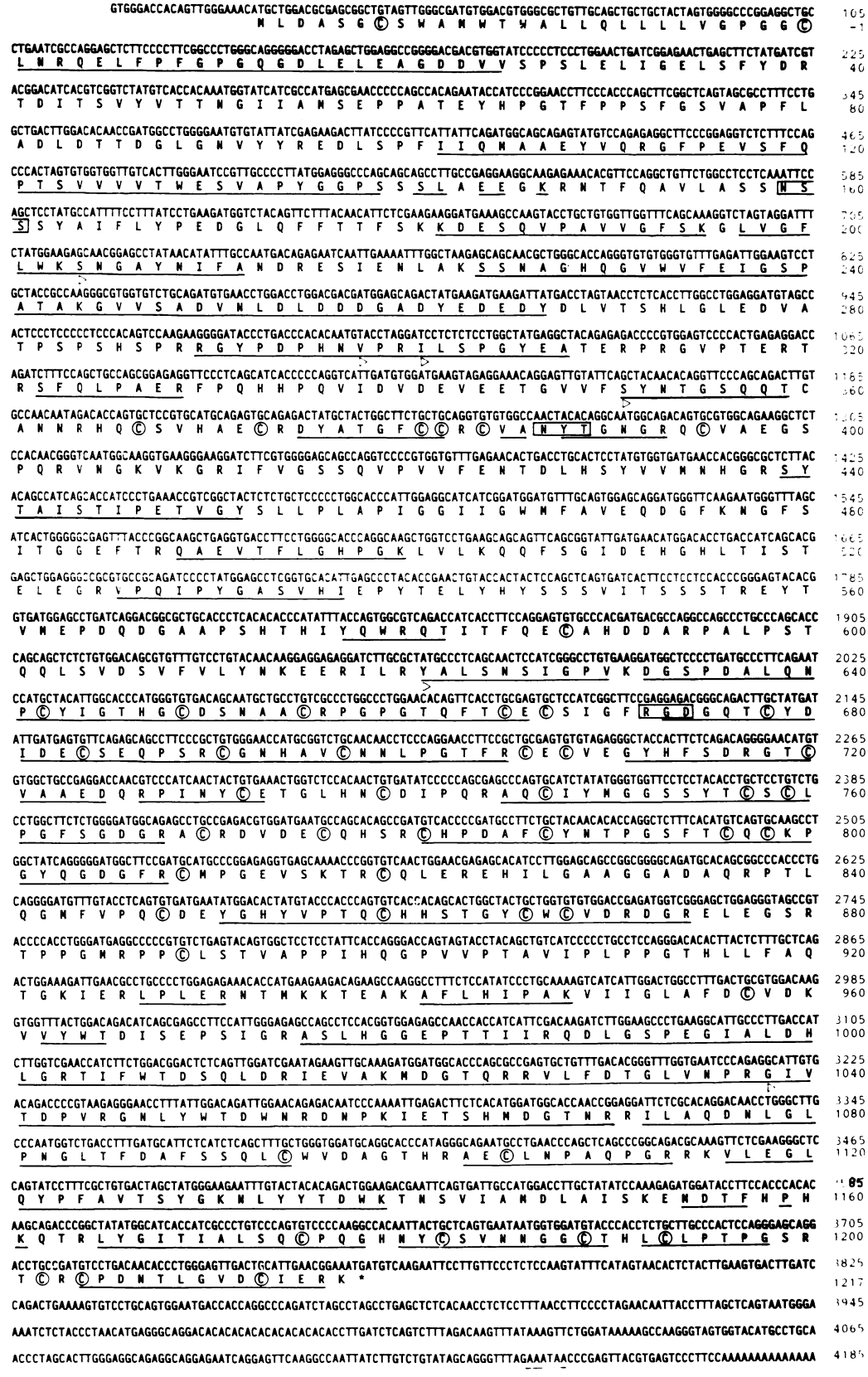

**Fig. 3.** Nucleotide and amino acid sequence of mouse nidogen. The nucleotide sequence was determined from cDNA clones shown in Figure 1. Underlined amino acid sequences were confirmed by Edman degradation of peptides isolated from various nidogen domains. Cysteine residues are encircled and potential *N*-linked carbohydrate acceptor sites and a RGD cell-binding site are shown in boxes. Open arrow heads mark the N-termini of endogenous nidogen fragments Nd-130 (position 204), Nd-100 (position 298), Nd-80 (position 351) and Nd-40 (position 621) described in previous studies (Dziadek *et al.*, 1985b; Paulsson *et al.*, 1986) and of the thrombin fragments Th-100 (position 301) and Th-22 (position 1038) (Mann *et al.*, 1988). A potential polyadenylation signal is underlined. Differences between the deduced amino acid sequence and results of Edman degradation (given in brackets) were noted for positions 140 (G), 143 (V), 146 (A) and 631 (R). They are all explained by single base substitutions.

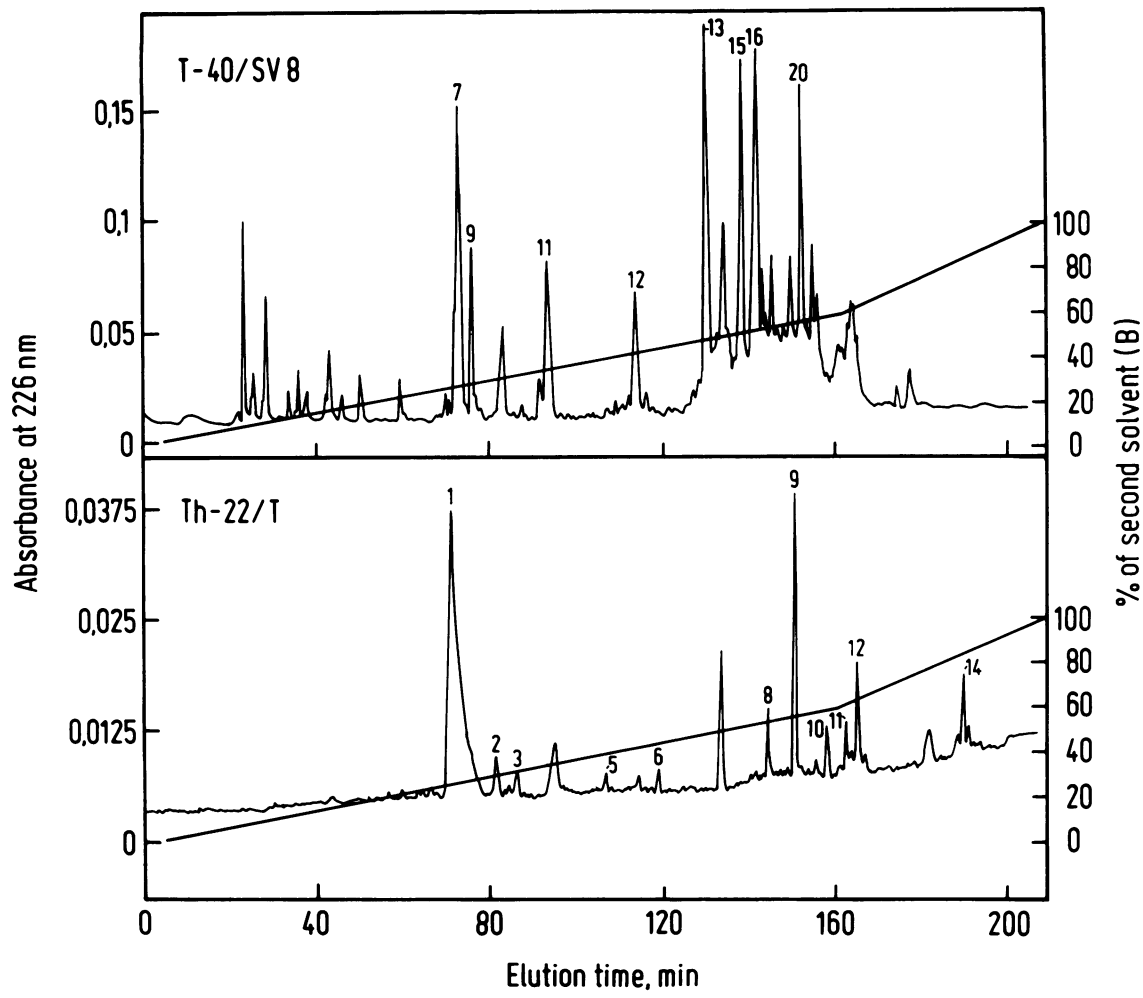


**Fig. 4.** HPLC profiles of tryptic peptides obtained from the rod-like domain (T-40/SV8) and a portion of the C-terminal globular domain (Th-22/T) of nidogen. The numbered peaks contained peptides identified by Edman degradation. Peptides in the upper panel were additionally cleaved with SV8 protease prior to separation.

to those of intact nidogen (Figure 5). This fragment corresponds to the rod-like domain and portions of the N-terminal globule of nidogen. Cell attachment to nidogen could also be inhibited by synthetic peptides. It was reduced to a 20% level by low concentrations of GRGDS (10 μg/ml) while related peptides RGES and AGDV required more than a 10- to 50-fold higher concentration to obtain the same effect.

## Discussion

The cDNA clones described here allowed us to determine the entire coding sequence of mouse nidogen mRNA with an ORF of 3735 nt which encode some 1217 amino acid residues of nidogen and a typical 28-residue signal peptide. Edman degradation of nidogen peptides was used to identify the N- and C-terminal ends of the protein and to confirm ~50% of the deduced amino acid sequence. The 28 amino acid residues predicted to precede the N-terminus show the characteristic features of a signal peptide (von Heijne, 1986). Whether the signal peptide in fact starts at a second Met at position −18 remains open. According to the scanning model of eukaryotic ribosome initiation (Kozak, 1984), the first ATG codon is more likely to be used as an initiator. In addition, it is in a better context for initiation than the second

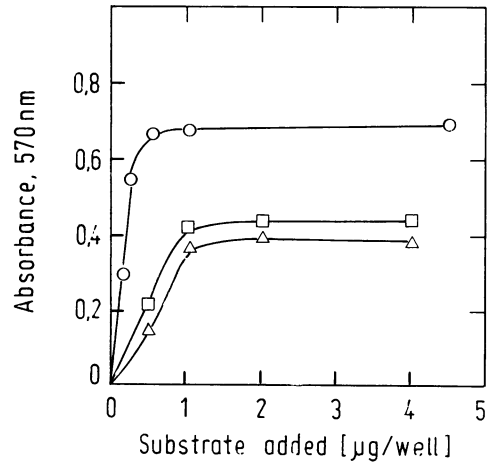


**Fig. 5.** Cell attachment of osteosarcoma cells SAOS-2 to nidogen. Substrates used were the laminin−nidogen complex (○), nidogen (△) and the nidogen fragment Nd-80 (□). Attached cells were quantified by a colorimetric assay.

ATG (Kozak, 1987) and the second ATG is not conserved in the human nidogen sequence (our unpublished observation).

Two mouse mRNA transcripts were recognized by the nidogen cDNA probe, a major one of 6 kb and a minor one

of 4 kb. These two mRNAs may differ in the coding or non-coding region. Several observations seem to favor the latter possibility. The ORF identified in the overlapping cDNA clones is of sufficient size to encode the entire nidogen sequence (see below). In addition, the 3′ untranslated region is terminated by a short poly(A) tail preceded by a variant of the consensus polyadenylation signal (AATAAA versus AAATAA; see Swimmer and Shenk, 1985) in two independent clones (N-2 and N-3). This suggests that these two clones are derived from the shorter mRNA species. Whether the longer mRNA species contains additional sequences only in the 3′ untranslated region remains to be determined.

The amino acid sequence of nidogen together with ~5% carbohydrate predicts a mol. wt of 141 kd which is in good agreement with a value of 148 kd determined by ultracentrifugation (Paulsson *et al.*, 1986). Furthermore, all the peptide sequences determined here and in previous studies were found in the coding region of the cDNA clones. The data also agree well with the dumb-bell model of nidogen proposed from electron microscopical observations (Paulsson *et al.*, 1986, 1987). They show that the 641 residue long N-terminal sequence forms the large globular domain I, followed by 248 residues in cysteine-rich repeats contributing the rod-like structure (domain II) and a smaller (328 residues) C-terminal globular domain III (Figure 6). Amino-terminal sequences described for nidogen and fragments produced by endogenous proteases during its extraction (Dziadek *et al.*, 1985b; Paulsson *et al.*, 1986) were all found in the N-terminal domain I with that of fragment Nd-40 being in front of domain II. This confirms predictions on the localization of these fragments based on electron microscopical and immunological studies (Mann *et al.*, 1988; Paulsson *et al.*, 1986). It also suggests that the bonds readily cleaved by endogenous proteases (Dziadek *et al.*, 1985a) are located in flexible segments of the N-terminal globule giving rise to three globular subdomains Ia, b and d and one EGF-like repeat Ic (Figure 6). Three subdomains can also be distinguished in the C-terminal globule, namely a hydrophobic and proline-rich segment IIIa close to the rod (positions 890–956), a globular structure IIIb with two putative disulfide bonds in the center followed by an EGF-like repeat IIIc directly at the C-terminus.

The rod-like domain of nidogen is composed of five consecutive EGF-like repeats (numbered IIa–IIe) which show internal homology in the positions of the six cysteines and several other residues (Figure 7) suggesting that they may have evolved by exon duplication. These segments apparently serve a structural purpose by separating the globular domains of nidogen by a 17 nm long, flexible link (Paulsson *et al.*, 1987) similarly to analogous repeats found in the rod-like portions of the short arms of laminin B1 and B2 chains (Sasaki *et al.*, 1987; Sasaki and Yamada, 1987). However, the latter structures are characterized by an eight rather than six cysteine residue repeat. In addition, such EGF-like repeats could express biological functions as indicated by a putative RGD cell-binding sequence being present between Cys-5 and Cys-6 of nidogen domain IIa (Figures 3 and 7). EGF-like repeats have been found in quite a few proteins and are commonly divided based on homologies into three different classes (Appella *et al.*, 1988). The EGF-like repeats of nidogen are similar to those in classes II or III which also include the EGF precursor, LDL receptor and several coagulation proteins. A particularly

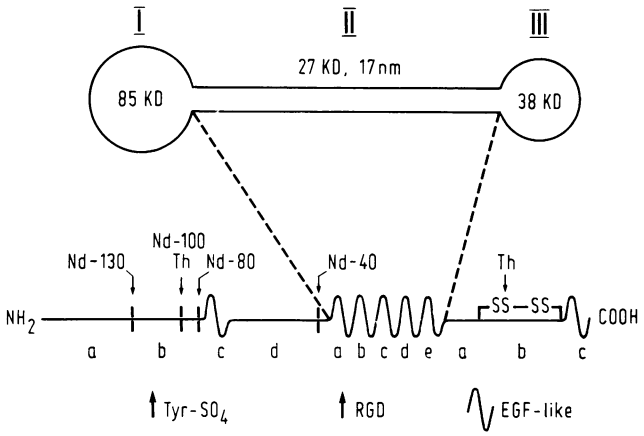


**Fig. 6.** Correlation between the amino acid sequence and domain structure of nidogen. The model of three major domains (I–III) is based on electron microscopy and ultracentrifugation (Paulsson *et al.*, 1986, 1987) and related to the sequence. Vertical black bars denote the start of various large nidogen fragments (Nd-130, Nd-100, Nd-80, Nd-40) which arise by endogenous proteolysis (Paulsson *et al.*, 1986), suggesting several subdomains within the large N-terminal globule (I). These and other subdomains are denoted by small letters. Five EGF-like repeats constitute the rod with a further repeat being present within domain I and at the C-terminus. Th denotes major sites of thrombin cleavage. Full arrows indicate positions of two potential tyrosine sulfate acceptor sites and of a single RGD sequence with the potential for cell-binding. SS denotes pairs of cysteine residues (besides those in the EGF-like repeats) possibly involved in intradomain disulfide bridges.

striking identity (50%) is observed between an EGF repeat of thyroglobulin (Malthiery and Lissitzky, 1987) and nidogen domain IIe which has unusually long inserts between Cys-1 and 2 and Cys-5 and 6 (Figure 7). The EGF-like repeat Ic is also atypical containing seven instead of six cysteines (Figure 7) which suggests one free thiol group or binding to the single cysteine (position 588) in domain Id. Odd numbers of cysteines in such repeats are rather uncommon but were found in the $\beta$ subunit of the fibronectin receptor (Argraves *et al.*, 1987).

Another homology with ~25% identical residues is found between nidogen domain III (positions 890–1209), EGF precursor (positions 487–775; Scott *et al.*, 1983) and LDL receptor (positions 372–703; Yamamoto *et al.*, 1983). However, this only becomes apparent by introducing several gaps and does not match the cysteine residues of subdomain IIIb (not shown). It remains therefore doubtful whether this similarity indicates any structural or functional relationship. Our sequence data also shed some light on the possible identity of nidogen (Paulsson *et al.*, 1985) and entactin (Carlin *et al.*, 1981, 1983). The amino acid sequence deduced from a 1.3 kb entactin cDNA clone (Durkin *et al.*, 1987) corresponds with a few exceptions to the nidogen sequence position 63–385 but is followed by 0.35 kb of a completely different sequence. Since the segment at variance is supported by three nidogen peptide sequences (starting at positions 385, 439 and 489, respectively, see Figure 3) it indicates that the entactin clone is a rearranged product. It nevertheless also indicates that nidogen and entactin are identical.

The nidogen sequence contains two Tyr residues (positions 262 and 267) in the sequence DDDGADYEDEDYD, which due to the acidic neighbouring residues represent typical acceptor sites for *O*-sulfation (Huttner and Bäuerle, 1988).

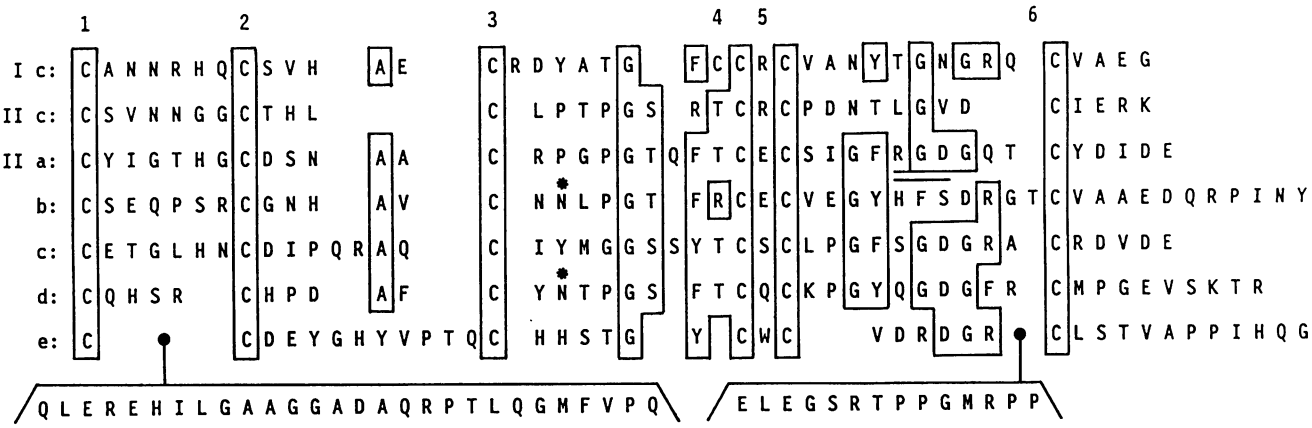


**Fig. 7.** Comparison of seven EGF-like repeats of nidogen with regions of high homology shown in boxes. Numbering of cysteines on top follows those for EGF. Sequences shown correspond to positions 360–399 (Ic), 1182–1217 (IIIc) and 642–900 (IIa–e) in Figure 3. Two Asn residues within consensus sequences for β-hydroxylation (Rees *et al.*, 1988) are indicated by an asterisk. An RGD sequence is marked by a double line.

This is of some interest since entactin was originally identified by sulfate labeling (Carlin *et al.*, 1981; Hogan *et al.*, 1982). Other studies with nidogen showed the label to be present on tyrosine possibly involving two residues in the N-terminal thrombin fragment Th-50 (Paulsson *et al.*, 1985; and unpublished). This is in good agreement with the sequence data and allows a precise identification of the sulfate acceptor sites. Nidogen also possesses two NXT/NXS sites (positions 159 and 387; Figure 3), putative acceptors for *N*-linked oligosaccharides. With a 2.0% glucosamine content (Paulsson *et al.*, 1986) it indicates that all acceptor sites are occupied by complex types of oligosaccharide side chains and is in agreement with the failure to identify Asn residue 387 by Edman degradation. About 0.9% galactosamine in nidogen also suggests the presence of six or seven *O*-linked oligosaccharides which have not been assigned so far to specific sequence positions.

A further post-translational modification of nidogen may include the β-hydroxylation of two Asn residues present in the EGF-like repeats IIb and IId (Figure 7). Hydroxylated Asn or Asp residues are often found in EGF-like domains of various calcium-binding proteins (i.e. coagulation factors, complement components) and are surrounded at some variable distance by the consensus sequences $DX^{D}_{N}XC$ and $^{F}_{Y}XC$ (Rees *et al.*, 1988). These consensus sequences are also present in nidogen (Figure 7) and small amounts of β-hydroxyaspartate have been recently detected in nidogen hydrolysates (S.Stenflo, personal communication). This suggests that nidogen is also a calcium-binding protein as found recently for other basement membrane components such as laminin (Paulsson, 1988) and BM-40 (Engel *et al.*, 1987).

The identification of an RGD sequence within the EGF-like repeat IIa provided the first strong indication that nidogen like laminin (Kleinman *et al.*, 1985; Martin and Timpl, 1987) and collagen IV (Aumailley and Timpl, 1986) could belong to the potential cell-binding basement membrane proteins. This was confirmed in cell attachment studies, as a major binding site was located to domains Ic and Id plus II of nidogen (fragment Nd-80). Cell binding was also inhibited by synthetic RGD peptides while control peptides differing at the Arg or Asp position were less active There are permutations of the RGD motif including DRG, DGR and RDG in other EGF-like repeats in nidogen (Figure 7). This is of some interest since the inverted cell-binding sequences DGR has been shown to interfere with cell-binding to laminin and fibronectin (Yamada and Kennedy, 1987). The evaluation of the biological importance of these sequences will require cell-binding studies with various relevant synthetic peptides from nidogen domain II. Our observations also suggest that nidogen binds to cellular receptors of the integrin family (Hynes, 1987; Ruoslahti, 1988). RGD-dependent integrin binding has for example been shown for fibronectin (Pytela *et al.*, 1985), vitronectin (Pytela *et al.*, 1986) and collagen I (Dedhar *et al.*, 1987). Whether nidogen can bind to one or all of these receptors or requires a specific integrin receptor remains to be determined.

Nidogen binds strongly to laminin via its C-terminal domain (Paulsson *et al.*, 1987; Mann *et al.*, 1988). Laminin is also known to possess two high affinity cell-binding sites (Aumailley *et al.*, 1987; Goodman *et al.*, 1987). Since laminin and nidogen seem to be complexed in a variety of tissues (Dziadek and Timpl, 1985), this interaction should increase the potential for cellular recognition within a narrow topological array. The C-terminal domain III of nidogen also has the potential to bind to two different sites in the triple helical domain of collagen IV although with apparently lower affinity compared to the binding of laminin (M.Aumailley *et al.*, in preparation). The precise identification of the binding sequences of nidogen for laminin and collagen IV seems feasible based on the information provided here and may then be exploited in further functional studies. This should assist in a better understanding of the obviously complex role played by nidogen in basement membranes.

## Materials and methods

***Isolation of nidogen peptides and amino acid sequencing***

Large fragments of nidogen used for sequence analysis included the endogenous fragment Nd-100, the thrombin fragments Th-100, Th-50 and Th-22 and the tryptic fragment T-40 (Paulsson *et al.*, 1986; Mann *et al.*, 1988). They were cleaved either prior to or after reduction of disulfide bonds with TPCK-trypsin (Mann *et al.*, 1988). Fragment T-40 which was released from Nd-100 prior to reduction was reduced and alkylated in 6 M guanidine–HCl (Paulsson *et al.*, 1986) and subjected to a second digestion by SV8 protease (enzyme–substrate ratio 1:50, 8 h at 20°C). The various peptide mixtures were separated on a reversed phase column (RP 318, C18-coated, BioRad Laboratories, Munich) which was equilibrated in 0.1% trifluoroacetic acid and eluted with a gradient of acetonitrile (Chu *et al.*, 1987). Digests of non-reduced Nd-100 and Th-100 were initially separated

into large and smaller fragments by chromatography on a Sephadex G-150 column equilibrated in 0.2 M ammonium bicarbonate.

Edman degradations were performed with 0.01–1 nmol peptide samples on a gas phase sequencer or a spinning-cup liquid-phase sequencer as previously described (Chu *et al.*, 1987; Hartl *et al.*, 1988). Amino acid sequences were screened for internal repeats and homologies to other proteins by the programs FASTP (Lipman and Pearson, 1985) DOTMATRIX (PSQ), SEQHP and SEQDP.

### *Isolation of cDNA clones and Northern blots*

Total RNA was prepared from mouse EHS tumor, differentiated F9 teratocarcinoma cells and human hepatoma cells (Hep G2, provided by B.B.Knowles and D.P.Alden, The Wistar Institute) using guanidine thiocyanate extraction and CsCl centrifugation (Chirgwin *et al.*, 1979). The F9 cells were induced to differentiate by retinoic acid and cAMP as previously described (Sasaki *et al.*, 1987). Poly(A)$^+$ RNA was selected on an oligo(dT)–cellulose column. Three λgt11 cDNA libraries were prepared from EHS tumor or F9 cell poly(A)$^+$ RNA using either oligo(dT)$_{12-18}$ or mixed hexanucleotides (Pharmacia, Piscataway, NJ) as primers by standard procedures (Gubler and Hoffman, 1983; Huynh *et al.* 1985). An additional cDNA library in the λZAP vector (Stratagene, La Jolla, CA) was prepared from EHS tumor poly(A)$^+$ RNA as described above except that *Eco*RI methylase was used to protect the *Eco*RI sites in the cDNA.

Two expression libraries prepared in the λgt11 vector from mouse EHS tumor mRNA were screened with rabbit antisera against mouse nidogen or nidogen fragments (Paulsson *et al.*, 1986) following established procedures (Young and Davies, 1983; Chu *et al.*, 1987). Four clones (insert size 0.5–1.5 kb) were eventually obtained from $2 \times 10^6$ plaques, were plaque-purified and shown by sequence analysis to correspond to nidogen. The largest insert was $^{32}$P-labelled by nick-translation (Rigby *et al.*, 1977), and used as a probe to screen the cDNA libraries from F9 mRNA ($1 \times 10^6$ plaques) and from EHS tumor mRNA in λZAP ($5 \times 10^5$ plaques from unamplified library) as described (Benton and Davis, 1977). Inserts from λgt11 libraries were isolated from agarose gels (Vogelstein and Gillespie, 1979) and subcloned into the *Eco*RI site of the Bluescript plasmid (Stratagene, La Jolla, CA). Clones from the λZAP library were infected with single-stranded helper phage R408 to convert the clones from phage vector to Bluescript vector according to the protocol provided by Stratagene.

Poly(A)$^+$ RNA was electrophoresed on a 1% agarose gel containing 6% formaldehyde, transferred to nitrocellulose (Thomas, 1980), and probed with inserts from positive clones. The RNA filters were hybridized and washed as previously described (Chu *et al.*, 1987).

### *Nucleotide sequencing*

Positive inserts in Bluescript plasmids were initially sequenced from both 5′ and 3′ directions by dideoxy chain termination sequencing of double-stranded DNA (Chen and Seeburg, 1985), using several primers flanking the cloning sites for the vector (Stratagene, La Jolla, CA). To determine the complete nucleotide sequence of the cDNA, the inserts were subcloned into various restriction sites of the M13 vector and single-stranded DNA templates were sequenced by use of M13 universal primers or specific primers synthesized based on available cDNA sequences. The region covered by clone 6c was determined on average five times using the shotgun sequencing strategy (Staden, 1982; Deininger, 1983). Most sequencing was performed with modified T7 polymerase (Sequenase, USB, Cleveland, OH) in dideoxy sequencing reaction mixtures containing (α-S)thio-dATP (New England Nuclear, Boston, MA), following the Sequenase kit manual (USB). Some sequences were determined with four fluorescent M13 primers following sequencing reaction conditions using Sequenase as suggested by Applied Biosystems (Foster City, CA). Nonradioactive samples were analyzed using an automated DNA sequencer (Applied Biosystems, Model 370A).

### *Cell attachment assays*

Tissue culture plastic multiwell plates (Costar, Cambridge, MA) were coated overnight at 4°C (Aumailley *et al.*, 1987) with either the laminin–nidogen complex, nidogen separated from the complex by dissociation with 2 M guanidine–HCl (Paulsson *et al.*, 1987) or with nidogen fragments. Attachment assays were then carried out with cell suspensions in serum-free medium for 3 h in the presence of cycloheximide (25 μg/ml) in a humified incubator at 37°C (Aumailley and Timpl, 1986). Attached cells were analyzed after staining with 0.1% crystal violet (Gillies *et al.*, 1986) in an ELISA reader (Dynatech MR 6000) at 570 nm. For inhibition assays cell suspensions were mixed with synthetic peptides dissolved in culture medium at different concentrations prior to the addition to the wells. Peptides were purchased from Promega, Switzerland, and Peninsula, St Helens.

## Acknowledgements

We thank Mrs Vera van Delden, Hildegard Reiter and Loretta Renkart for skilled technical assistance. The study was supported by grants of the Fritz Thyssen Stiftung, Deutsche Forschungsgemeinschaft and NIH AR-38912, AR-38923 and AR-38188.

## References

Appella,E., Weber,I.T. and Blasi,F. (1988) *FEBS Lett.*, **231**, 1–4.
Argraves,W.S., Suzuki,S., Arai,H., Thompson,K., Pierschbacher,M.D. and Ruoslahti,E. (1987) *J. Cell Biol.*, **105**, 1183–1190.
Aumailley,M. and Timpl,R. (1986) *J. Cell Biol.*, **103**, 1569–1575.
Aumailley,M., Nurcombe,V., Edgar,D., Paulsson,M. and Timpl,R. (1987) *J. Biol. Chem.*, **262**, 11532–11538.
Benton,W.D. and Davis,R.W. (1977) *Science*, **196**, 180–182.
Carlin,B., Jaffe,R., Bender,B. and Chung,A.E. (1981) *J. Biol. Chem.*, **256**, 5209–5214.
Carlin,B.E., Durkin,M.E., Bender,B., Jaffe,R. and Chung,A.E. (1983) *J. Biol. Chem.*, **258**, 7729–7737.
Chen,E.Y. and Seeburg,P.H. (1985) *DNA*, **4**, 165–170.
Chirgwin,J.M., Przybyla,A.E., MacDonald,R.J. and Rutter,W.J. (1979) *Biochemistry*, **18**, 5294–5299.
Chu,M.-L., Mann,K., Deutzmann,R., Pribula-Conway,D., Hsu-Chen, C.-C., Bernard,M.P. and Timpl,R. (1987) *Eur. J. Biochem.*, **168**, 309–317.
Dedhar,S., Ruoslahti,E. and Pierschbacher,M.D. (1987) *J. Cell Biol.*, **104**, 585–593.
Deininger,P.G. (1983) *Anal. Biochem.*, **129**, 216–223.
Durkin,M.E., Carlin,B.E., Vergnes,J., Bartos,B., Marlie,J. and Chung,A.E. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 1570–1574.
Dziadek,M. and Timpl,R. (1985) *Dev. Biol.*, **111**, 372–382.
Dziadek,M., Paulsson,M. and Timpl,R. (1985a) *EMBO J.*, **4**, 2513–2518.
Dziadek,M., Paulsson,M., Deutzmann,R., Timpl,R., Weber,S. and Engel,J. (1985b) In Shibata,S. (ed.), *Basement Membranes*. Elsevier, Amsterdam, pp. 13–23.
Engel,J., Taylor,M., Paulsson,M., Sage,H. and Hogan,B.L.M. (1987) *Biochemistry*, **26**, 6958–6965.
Gillies,R.J., Didier,N. and Denton,M. (1986) *Anal. Biochem.*, **159**, 109–113.
Goodman,S.L., Deutzmann,R. and von der Mark,K. (1987) *J. Cell Biol.*, **105**, 589–598.
Gubler,U. and Hoffman,B.J. (1983) *Gene*, **25**, 263–269.
Hartl,L., Oberbäumer,I. and Deutzmann,R. (1988) *Eur. J. Biochem.*, **173**, 629–635.
Hogan,B.L.M., Taylor,A., Kurkinen,M. and Couchman,J.R. (1982) *J. Cell Biol.*, **95**, 197–204.
Huttner,W.B. and Bäuerle,P.A. (1988) *Modern Cell Biol.*, **6**, 97–140.
Huynh,T.V., Young,R.A. and Davis,R.W. (1985) In Glover,D.M. (ed.), *DNA Cloning*. IRL Press, Oxford, Vol. I, pp. 49–78.
Hynes,R.O. (1987) *Cell*, **48**, 549–554.
Kleinman,H.K., Cannon,F.B., Laurie,G.W., Hassell,J.R., Aumailley,M., Terranova,V.P., Martin,G.R. and Dubois-Dalcq,M. (1985) *J. Cell. Biochem.*, **27**, 317–325.
Kozak,M. (1984) *Nucleic Acids Res.*, **12**, 3873–3893.
Kozak,M. (1987) *Nucleic Acids Res.*, **15**, 8125–8147.
Lipman,D.J. and Pearson,W.R. (1985) *Science*, **227**, 1435–1441.
Malthiery,Y. and Lissitzky,S. (1987) *Eur. J. Biochem.*, **165**, 491–498.
Mann,K., Deutzmann,R. and Timpl,R. (1988) *Eur. J. Biochem.*, in press.
Martin,G.R. and Timpl,R. (1987) *Annu. Rev. Cell Biol.*, **3**, 57–85.
Paulsson,M. (1988) *J. Biol. Chem.*, **263**, 5425–5430.
Paulsson,M., Dziadek,M., Suchanek,C., Huttner,W.B. and Timpl,R. (1985) *Biochem. J.*, **231**, 571–579.
Paulsson,M., Deutzmann,R., Dziadek,M., Nowack,H., Timpl,R., Weber,S. and Engel,J. (1986) *Eur. J. Biochem.*, **156**, 467–478.
Paulsson,M., Aumailley,M., Deutzmann,R., Timpl,R., Beck,K. and Engel,J. (1987) *Eur. J. Biochem.*, **166**, 11–19.
Pytela,R., Pierschbacher,M.D. and Ruoslahti,E. (1985) *Cell*, **40**, 191–198.
Pytela,R., Pierschbacher,M.D., Ginsberg,M.H., Plow,E.F. and Ruoslahti,E. (1986) *Science*, **231**, 1559–1562.
Rees,D.J.G., Jones,I.M., Handford,P.A., Walter,S.J., Esnouf,M.P., Smith,K.J. and Brownlee,G.G. (1988) *EMBO J.*, **7**, 2053–2061.
Rigby,P.W.I., Dieckmann,M., Rhodes,C. and Berg,P. (1977) *J. Mol. Biol.*, **113**, 237–251.
Ruoslahti,E. (1988) *Annu. Rev. Biochem.*, **57**, 375–413.
Sasaki,M. and Yamada,Y. (1987) *J. Biol. Chem.*, **262**, 17111–17117.

Sasaki,M., Kato,S., Kohno,K., Martin,G.R. and Yamada,Y. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 935–939.
Scott,J., Urdea,M., Quiroga,M., Sanchez-Pescador,R., Fong,N., Selby,M., Rutter,W.J. and Bell,G.I. (1983) *Science*, **221**, 236–240.
Staden,R. (1982) *Nucleic Acids Res.*, **10**, 2951–2961.
Swimmer,C. and Shenk,T.C. (1984) *Nucleic Acids Res.*, **13**, 8053–8063.
Thomas,P.S. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5201–5206.
Timpl,R. and Dziadek,M. (1986) *Int. Rev. Exp. Pathol.*, **29**, 1–112.
Vogelstein,B. and Gillespie,D. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 615–619.
von Heijne,G. (1986) *Nucleic Acids Res.*, **14**, 4683–4690.
Yamada,K.M. and Kennedy,D.W. (1987) *J. Cell Physiol.*, **130**, 21–28.
Yamamoto,T., Davis,C.G., Brown,M.S., Schneider,W.J., Casey,M.L., Goldstein,J.L. and Russell,D.W. (1984) *Cell*, **39**, 27–38.
Young,R.A. and Davis,R.W. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 1194–1198.

*Received on September 26, 1988*