

# Cloning and Expression of cDNA Encoding a Neural Calcium-dependent Cell Adhesion Molecule: Its Identity in the Cadherin Gene Family

Kohei Hatta, Akinao Nose, Akira Nagafuchi, and Masatoshi Takeichi

Department of Biophysics, Faculty of Science, Kyoto University, Kitashirakawa, Sakyo-ku, Kyoto 606, Japan

**Abstract.** The neural cadherin (N-cadherin) is a  $\text{Ca}^{2+}$ -dependent cell-cell adhesion molecule detected in neural tissues as well as in non-neural tissues. We report here the nucleotide sequence of the chicken N-cadherin cDNA and the deduced amino acid sequence. The sequence data suggest that N-cadherin has one transmembrane domain which divides the molecule into an extracellular and a cytoplasmic domain; the extracellular domain contains internal repeats of characteristic sequences. When the N-cadherin cDNA connected with virus promoters was transfected into L

cells which have no endogenous N-cadherin, the transformants acquired the N-cadherin-mediated aggregating property, indicating that the cloned cDNA contained all information necessary for the cell-cell binding action of this molecule. We then compared the primary structure of N-cadherin with that of other molecules defined as cadherin subclasses. The results showed that these molecules contain common amino acid sequences throughout their entire length, which confirms our hypothesis that cadherins make a gene family.

**T**HE  $\text{Ca}^{2+}$ -dependent cell-cell adhesion molecule detected in neural tissues has been defined as N-cadherin<sup>1</sup> (6, 7). This molecule appears to play a key role in constructing early embryonic neural tissues, since antibodies against N-cadherin tend to dissociate them (7).

N-cadherin is expressed not only in neural tissues but also in many nonneural tissues (6, 7, 8). In the development of embryos, the expression of N-cadherin begins in the mesoderm upon gastrulation, and subsequently occurs in the neural tube. Thereafter, the pattern of N-cadherin expression dynamically changes during development and its appearance or disappearance is strongly correlated with morphogenetic events such as segregation or association of cells. It was also observed that changes in the localization of N-cadherin on individual cells are associated with the rearrangement of cells from the mesenchyme to the epithelium or vice versa (4, 8).

We previously identified two other immunologically distinct  $\text{Ca}^{2+}$ -dependent cell-cell adhesion molecules, the epithelial cadherin (E-cadherin) (24) and the placental cadherin (P-cadherin) (16), each showing a unique tissue distribution pattern. Expression of these molecules is also closely associated with morphogenetic events in development. From these observations, we suggested that differential expression of the three cadherins plays an important role in the organization of the animal body (20).

1. *Abbreviations used in this paper:* E-cadherin, epithelial cadherin; L-CAM, liver cell adhesion molecule; N-cadherin, neural cadherin; P-cadherin, placental cadherin.

N-, E-, and P-cadherin are similar in their  $\text{Ca}^{2+}$ -sensitivity as well as in molecular mass, but they distinguish themselves in their immunological and cell-cell binding specificities (see 20 for review). Our recent studies suggested that E- and P-cadherin are genetically related to each other, since they show 58% identity in their amino acid sequences (17). We also found that N- and E-cadherin share common amino acid sequences at their  $\text{NH}_2$  termini, as revealed by the direct amino acid sequencing of purified peptides (19). All these data suggest that cadherins constitute a gene family.

In the present investigation, we cloned cDNA encoding N-cadherin from chicken neural tissues, and determined its nucleotide sequence. Comparison of the deduced amino acid sequence of N-cadherin with that of other cadherins confirmed our hypothesis that all these molecules have a common ancestor. We also succeeded in transfecting a mouse cell line not expressing this molecule with the N-cadherin cDNA. The transformed cells acquired N-cadherin-mediated cell-cell adhesion properties, providing the first positive evidence that N-cadherin is a cell-cell adhesion molecule.

## Materials and Methods

### Rabbit Antiserum against N-Cadherin

87-kD tryptic fragments of N-cadherin were purified from newly hatched chicken brains as previously described (19). 10–20  $\mu\text{g}$  of these molecules emulsified with Freund's complete adjuvant were injected subcutaneously into a rabbit. The rabbit was boosted three times at 10-d intervals with 10–20  $\mu\text{g}$  fragments emulsified with incomplete Freund's adjuvant. After the last

boost, antiserum was collected. Western blot analysis showed that the antiserum specifically reacts with N-cadherin peptides.

### cDNA Libraries

A randomly primed  $\lambda$ gt11 cDNA library was prepared from poly(A)<sup>+</sup>RNA isolated from brain and retina of 7.5-d-old chicken embryos by the standard method (10). An oligo(dT)-primed  $\lambda$ gt11 library was prepared from poly(A)<sup>+</sup>RNA obtained from brain and retina of 6.5-d-old chicken embryos.

The randomly primed  $\lambda$ gt11 library was screened with the rabbit antiserum against N-cadherin. Approximately 10<sup>6</sup> recombinants were screened. Using one positive clone,  $\lambda$ N1, obtained in the first screening, more clones,  $\lambda$ N2,  $\lambda$ N3, and  $\lambda$ N4, were isolated from these libraries.

### Northern and Southern Blot Analysis

For Northern blot analysis, total RNA isolated from tissues by the guanidine method (14) was resolved in a 1% agarose/formaldehyde (2  $\mu$ g RNA/lane), transferred to a nitrocellulose filter, and hybridized at 42°C in the buffer containing 50% formamide with <sup>32</sup>P-labeled  $\lambda$ N1 insert.

For Southern hybridization, DNA was isolated from a whole 11-d-old chicken embryo, digested with Eco RI, Sac I, or Pst I, and resolved in a 0.7% agarose gel (10  $\mu$ g DNA/lane). DNA was then transferred to a nitrocellulose filter and hybridized under the same conditions as for the Northern blot analysis.

### Sequence Analysis

After subcloning of the DNA restriction fragments into pUC19 or pUC18 cloning sites, nucleic acid sequences were determined by the dideoxynucleotide chain termination method of Sanger et al. (18) using 7-deazaguanine instead of guanine. In most cases, overlapping sequence data were obtained for both strands.

### Plasmid Construction

pST-Ncad was constructed from pSTneo (12), replacing the neoresistant gene with the  $\lambda$ N2 insert, as follows: DNA of  $\lambda$ N2 was partially digested with Eco RI, and the full length insert which has Eco RI linkers at both sides was isolated. This insert was recloned into the Eco RI site of SK poly-linker of the Blue script M13+ (Stratagene Cloning Systems, San Diego, CA) in an inverted orientation. The insert was reisolated by double digestion with Eco RV and Sma I. pSTneo was double digested with Bgl II and Sma I, and the vector containing promoters of SV40 and HSV thymidine kinase was isolated. The vector was blunt ended, and was ligated with the  $\lambda$ N2 insert.

### Transfection of Cells

L cells were grown in a 1:1 mixture of Dulbecco's modified Eagle's minimal essential medium and Ham's F12 supplemented with 10% FCS (DH10) in Petri dishes (No. 3001; Falcon Plastics, Cockeysville, MD). 1  $\mu$ g of pST-Ncad and 0.1  $\mu$ g of pSTneo were coprecipitated by the calcium phosphate method (23), and added to L-cell cultures. About one day after transfection, cells were transferred into larger dishes (No. 3003; Falcon Plastics) and cul-

tured in DH10 containing 0.4 mg/ml G418. After ~1 wk, colonies were isolated, and tested for immunoreactivity to the monoclonal antibody NCD-2 (6), and then positive cells were recloned.

### Cell Aggregation

Cadherin-mediated cell aggregation was assayed as described (22). Briefly, cells were treated with 0.01% crystallized trypsin in the presence of 1 mM CaCl<sub>2</sub> at 37°C for 20 min, and then washed with a Ca<sup>2+</sup>-free saline to obtain single cell suspensions. Cells suspended with or without 1 mM CaCl<sub>2</sub> were placed in each well of a plate (No. 3047; Falcon Plastics; 10<sup>5</sup> cells/0.5 ml medium/well) and incubated to allow aggregation for 20 min at 37°C on a gyratory shaker rotated at 80 rpm. To test the effect of the NCD-2 antibody, 60  $\mu$ g/ml of antibody were added to the medium. The extent of aggregation was represented by an index  $N_{20}/N_0$ , where  $N_{20}$  is the number of cell clusters at an incubation time of 20 min, and  $N_0$  is the number of total cells.

### Results

#### Molecular Cloning and Characterization of Chicken N-Cadherin cDNA

The  $\lambda$ gt11 expression cDNA library was screened with the rabbit antiserum against N-cadherin. One positive clone designated as  $\lambda$ N1 was isolated from about 10<sup>6</sup> independent clones. Fusion proteins derived from the  $\lambda$ N1 reacted with the NCD-2 monoclonal antibody which specifically recognizes N-cadherin (6). Nucleotide sequencing showed that  $\lambda$ N1 contains a sequence corresponding to the NH<sub>2</sub>-terminal amino acid sequence of a tryptic fragment of the mature N-cadherin, which was determined using the gas-phase amino acid sequencer (19). Northern hybridization analysis showed that the tissue distribution of RNA hybridized with the  $\lambda$ N1 is identical to that of the N-cadherin protein; the hybridized band was detected in brain and heart, but not in liver (Fig. 2 a). From these observations, we concluded that  $\lambda$ N1 encodes N-cadherin.

Using  $\lambda$ N1 as a probe, we screened the above library and also another  $\lambda$ gt11 cDNA library made with the oligo(dT) primer to cover the entire of the mRNA. Overlapping clones that were isolated are shown in Fig. 1.

Northern blot analysis of RNA derived from chicken embryonic brain and heart, using the  $\lambda$ N1 probe, gave a single band ~4.3 kb (Fig. 2 a). The total length of overlapping cDNA clones obtained is 3,221 bases, suggesting that they do not cover the entire length of the mRNA. However, according to the observations described below, the cloned

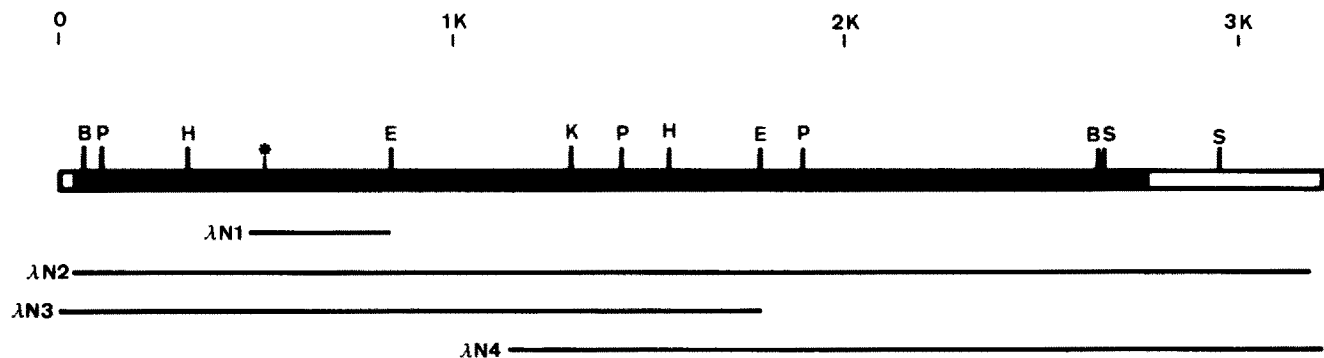
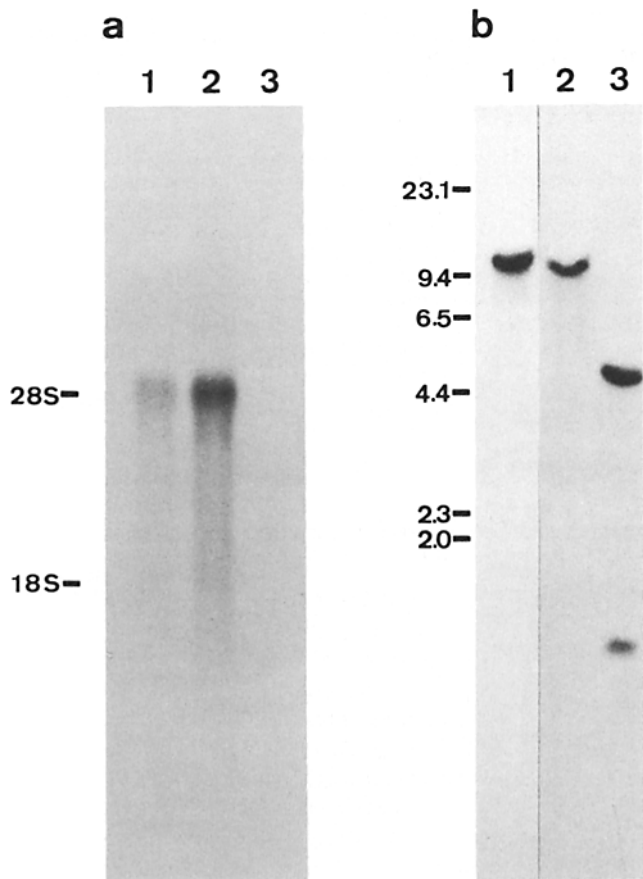


Figure 1. Restriction map of chicken N-cadherin cDNA. The solid box indicates the open reading frame. The position of the putative NH<sub>2</sub> terminus of the mature protein is shown by an asterisk. E, Eco RI; B, Bam HI; H, Hind III; K, Kpn I; P, Pst I; S, Sac I.



**Figure 2.** Northern and Southern analysis using  $\lambda$ N1 as a probe. (a) Total RNA extracted from brain (lane 1), heart (lane 2), and liver (lane 3) of 13-d-old chicken embryos by the guanidine method was resolved in a 1% agarose/formaldehyde gel (2  $\mu$ g RNA/lane). Positions of 28S and 18S ribosomal RNA are shown. (b) DNA isolated from a 11-d chicken embryo was digested with Eco RI (lane 1), Sac I (lane 2), and Pst I (lane 3), and resolved in a 0.7% agarose gel (10  $\mu$ g DNA/lane). Reference markers are shown in kilobases.

cDNAs probably contain all information necessary for the function of N-cadherin.

In Southern blot analysis using  $\lambda$ N1, the DNA digests with Eco RI and Sac I each gave a single band, and those with Pst I gave one major and one minor band, suggesting that the N-cadherin gene in the chicken genome is present in one or at most a few copies (Fig. 2 b).

#### **Nucleotide and Deduced Amino Acid Sequences of N-Cadherin**

The entire nucleotide sequence of overlapping cDNA clones and the putative amino acid sequence are shown in Fig. 3. The first ATG found at position 35 is a candidate for the initiation codon, since it is followed by a sequence encoding a stretch of hydrophobic amino acids that may function as a signal peptide. It may be possible that the initiation codon is localized further upstream, since no stop codon is found in the region preceding this ATG within the sequence presently determined. Transfection experiments described below, however, showed that the cDNAs obtained here contain

all information necessary for the expression of N-cadherin on the cell membrane.

The putative initiation codon is followed by an open reading frame of 2,736 bases and then by a TGA termination codon at position 2,771. This open reading frame encodes a polypeptide of 912 amino acids. The coding region is followed by 409 bases of 3'-untranslated sequence, which ends with a poly(A) addition sequence (ATTAAA) followed 18 bases later by a poly(A) stretch.

The NH<sub>2</sub>-terminal amino acid sequence of a tryptic fragment of N-cadherin (19) was found at position 165–173. Since the same sequence was also found at the NH<sub>2</sub> terminus of the mature form of liver cell adhesion molecule (L-CAM) (3), we tentatively assume that this site corresponds to the NH<sub>2</sub>-terminus of the mature N-cadherin. Under this assumption, N-cadherin is composed of 748 amino acids. Molecular mass calculated from this sequence is 81,800 D, which is much smaller than that determined by SDS-PAGE (6); this kind of discrepancy in the estimation of molecular mass was also found in other cadherins (5, 15, 17).

Hydrophobicity analysis according to the method of Kite and Doolittle (13) revealed that there is one hydrophobic stretch consisting of 32 uncharged amino acids starting at position 721, which shows characteristics of the transmembrane sequence, suggesting that N-cadherin is a transmembrane protein. The NH<sub>2</sub>-terminal side of N-cadherin probably corresponds to the extracellular domain, since the NCD-2 monoclonal antibody that is known to recognize an extracellular region of this protein reacted with the fusion protein derived from the  $\lambda$ N1 cDNA probe covering the NH<sub>2</sub>-terminal sequence.

The computer analysis using the IDEAS program (11) suggested that there is an internal repeat in the putative extracellular domain; the region from positions 198–304 showed 34% similarity to that from positions 309 to 418 (see Figs. 4 and 7). We also found that this domain contains internal repeats of many short sequences as shown in Fig. 4. It should be noted that negatively charged amino acids are highly conserved in the repeated sequences.

Potential N-linked glycosylation consensus sequences (N-X-T/S) were found at eight positions in the extracellular domain (Fig. 3). We found no significant sequence identity between N-cadherin and other proteins in the computer search using the NBRF database.

#### **Transfection of N-Cadherin cDNA into Mouse L Cells**

To learn whether the cloned N-cadherin cDNA encodes functional molecules, we introduced the N-cadherin cDNA that was connected with virus promoters (pST-Ncad) (Fig. 5 a) into mouse L cells, together with the plasmid containing a neoresistant gene (pSTneo). L cells do not express cadherins that react with antibodies to the chicken N-cadherin.

Out of several cotransformants of pST-Ncad and pSTneo, we selected two clones, designated NL1 and NL2. Immunoblot analysis showed that these two clones express a 132-kD peptide recognized by antibodies to N-cadherin, whose size is identical to that of N-cadherin expressed in the heart of newly hatched chickens (Fig. 5 b). Some smaller peptides were also detected, which are possibly products of degradation or misprocessing.

$\lambda$ N2  
 GGGGGCCGCCCGCCGCCGCCCTCCTCGCCTCCATGTGCCGGATAGCGGGAACGCCGCCGGATCCTGCCCGCTGGCGCTGATGTCTGCGGCCCTGCAGCAGGCACCCGATA  
 M C R I A G T P P R I L P P L A L M L L A A L Q Q A P I

119 AAAGCAACTTGTGAAGACATGTTGTGAAGATGGGATTTCTGAAGATGTCACAGTGCAGTCTGTCGAGGAGTGTACATGGAGGACAACTCTGCTCAATGTGAGGTTTCAAAGCTGC  
 29 K A T C E D M L C K M G F P E D V H S A V V S R S V H G G Q P L L N V R F Q S C

239 GATGAAAACAGAAAAATATACTTTGGAAGCAGTGAGCCAGAAGATTTTAGAGTAGTGAAGATGGTGTGGTATAATGCAGAGAGAAGCTTTCAACTTTTCCAGCAGAGCCACGGAGTTTGTGA  
 69 D E N R K I Y F G S S E P E D F R V G E D G V V Y A E R S F Q L S A E P T E F V

359 GTGTCTGCTCGAGACAAGGAACTCAGGAAGAATGGCAAATGAAGTGAAGTAAACCCCTGAACCGACATTCACAGGGGCCCTCAGAAAAGGACAAAAGAAAATGAAGACATCATATTT  
 109 V S A R D K E T Q E E W Q M K V K L T P E P A F T G A S E K D Q K K I E D I I F

479 CCATGGCAACAATAAGGACAGCCATCTGAAGAGACAGAAGAGACTGGGTTATCCCTCCAATCAACCTACCAGAAAATTCAGAGGACCTTTCTCAAGAATTAGTTAGGATT  
 149 P W Q Q Y K D S S H L K R Q K R D W V I P P I N L P E N S R G P F P Q E L V R I  
 .....

599 CGGTCTGATCGTGATAAAAGCCTTTCCTGCTACGGTACAGTGTGACTGGCCAGGAGCTGACCAACCTCCAACAGGAATCTTCATCATCAACCCCTCTCAGGACAGCTGTCTGTGACAAA  
 189 R S D R D K S L S L R Y S V T G P G A D Q P P T G I F I I N P I S G Q L S V T K

719 CCTTTAGATCGGGAGCAGATGCTTCTTTTCTGCTGAGAGCACATGCAGTGGATGTAATGGAAACAGGTGAAAAATCCTATTTGATATTGTGATTAACGTCATGATGAATGATAAC  
 229 P L D R E Q I A S F H L R A H A V D V N G N Q V E N P I D I V I N V I D M N D N

839 AGACCTGAATTTTGCATCAGGTTTGGAAATGGGACAGTTCCTGAAGGATCAAAGCCAGGAACCTATGTAATGACTGTTACTGCCATCGATGTGTGATGCCAATGCACAGAATGGGATG  
 269 R P E F L H Q V W N G T V P E G S K P G T Y V M T V T A I D A D D P N A Q N G M

959 CTGAGATACAGAATCTTGTACAGGCACCAAGCAGTCCCTCTCCCAACATGTTTACAATCAACAATGAGACTGGTGACATTTATCCGTCAGCAGTGGGCTCGACAGAGAAAATGACAA  
 309 L R Y R I L S Q A P S S P S P N M F T I N N E T G D I I T V A A G L D R E K V Q

1079 CAGTATACATTAATAATCAAGCTACGGACATGGAAGGAAACCAACATATGGTCTTTCAAACACAGCAACTGCTGTCATCACTGTGACAGATGTCAATGACAATCCTCCAGAGTTCACT  
 349 Q Y T L I I Q A T D M E G N P T Y G L S N T A T A V I T V T D V N D N P P E F T

1199 GCTATGACTTTTCTACGGTGAAGTACCAGAAAACAGAGTGGATGTCATAGTGGTAACTTAACAGTAAACAGATAAAGATCAGCCACACACGCTCGCTGGAATGCGAGGTACCAATGACA  
 389 A M T F Y G E V P E N R V D V I V A N L T V T D K D Q P H T P A W N A R Y Q M T

1319 GGGGAGACCCGACAGCCAGTTTACTACTCTGACCGATCCAATAGCAATGATGGTGGTAACTGTTGTCAAGCCATTGACTTTGAGACCAACAGGATGTTTGTACTTACTGTAGCT  
 429 G G D P T G Q F T I L T D P N S N D G L V T V V K P I D F E T N R M F V L T V A

1439 GCAGAAAATCAAGTCCCTTTGGCTAAGGGGATTCAGCATCTCCTCAGTCAACAGCAACCGTTCACATTACAGTCAATGATGTAATGAGAGTCCATATTTGTTCAAACCCCAAGCTT  
 469 A E N Q V P L A K G I Q H P P Q S T A T V S I T V I D V N E S P Y F V P N P K L

1559 GTACGTCAAGAAGAGGGCTACTTGTGCTAGCATGTGACAACTTTCAGTCTCGGACCCAGATCGTTACATGCAGCAAACTCTCTAAGGTACTCAAACCTTTCCGACCCGCAAC  
 509 V R Q E E G L L A G S M L T T F T A R D P D R Y M Q Q T S L R Y S K L S D P A N

1679 TGGTAAAAATGACCTGTAAATGGACAAAATAACAACACAGCTGTTTGGACAGAGAATCGATATATGTGAAAAAATATGTATAATGCAACTTTTCTGCTCTGATAATGGAATT  
 549 W L K I D P V N G Q I T T T A V L D R E S I Y V Q N N M Y N A T F L A S D N G I

1799 CCTCCAATGAGTGAAGTGGTACTTTCAGATATACTTGTGACATCAATGATAATGCTCCCAAGTGAACCCAAAAGAGCCACCCTGTGAAACACTGCAGCCTAATGCTATTAAC  
 589 P P M S G T G T L Q I Y L L D I N D N A P Q V N P K E A T T C E T L Q P N A I N

1919 ATCACTGCTAGACCCGACATGATCCAAATGCAGGCCATTTGCCCTTTGAGTGCCTGATTCACCTCCTAGTATTAAGAGGAATGGACCAATTGTCGAATTAGTGGTATCATGCC  
 629 I T A V D P D I D P N A G P F A F E L P D S P P S I K R N W T I V R I S G D H A

2039 CAGCTCTTTAAGGATCAGGTTCTCGGAGGCTGGTATCTATGATGTGCCATAGTAATACAGATTCCTGAAAATCCACATGCATCTAGCACTTCTGTGCTAAAAGTGAAGTTTGGCAA  
 669 Q L S L R I R F L E A G I Y D V P I V I T D S G N P H A S S T S V L K V K V C Q

2159 TGTGACATAAATGGGACTGTACTGATGTTGACCCGGATGTTGGCGCAGGACTGGGCACTGGTCCATCATGCAATTCCTGTTTATCATCATCTTACTCATTTTGTATGTTGATGTT  
 709 C D I N G D C T D V D R I V G A G L G T G A I I A I L L C I I I L L I L V L M F

2279 GTAGTATGGATGAAGCCCGCTGATAAGGAGCGTCAGGCCAAGCAGCTCTTAATTTGATCCAGAAGATGATGTGAGGACAACATCTGAAAATATGATGAAGAAGGTGGTGGAGAAGAAGAT  
 749 V V W M K R R D K E R Q A K Q L L I D P E D D V R D N I L K Y D E E G G G E E D

2399 CAGGATATGACTTGAGCCAGCTCCAGCAGCTGCACACTGTAGAACCAGCCATCAAACTGTTGGAATCAGACGCTTGTGATGAAAGGCCAATCCATGCAGAACCTCAGTATCCAGTC  
 789 Q D Y D L S Q L Q Q P D T V E P D A I K P V I R R L D E R P I H A E P Q Y P V

2519 AGATCAGCTGCTCCTCATCTCGGGGACATTTGGGACTTCATTAATGAGGGACTTAAAGCAGCCGACACAGCCCTACAGCCCGCCATCAGATTCCTCTTAGTCTTTGACTATGAAGGA  
 829 R S A A P H P G D I G D F I N E G L K A A D N D P T A P P Y D S L L V F D Y E G

2639 AGCGGCTCCACTGCTGCTCTTGAAGTCTCTTAATTCCTCAAGTAGCGGTGGTGGAGCAAGACTATGACTACCTAAATGACTGGGGCCACGTTTCAAGAACTTGTGACATGTACGGT  
 869 S G S T A G S L S L S S S S G G E Q D Y D Y L N D W G P R F K K L A D M Y G

2759 GGAGGTGATGACTGAACCTCAAGTGAACCTTTGTTTCTGGACAAGTACAACATTTCAACTGATATTTCCCAAAAAGCATACAGAAGCTAGGCTTTAACTTTGTAGTCTACTAGCACAACT  
 909 G G D D \*

2879 GCTTGTGAAGGCTTTGGCATAGGCTGCAACCAATTTGGGCTCCGTGGAATATCAGTATCCAACTGCTGTTTGGAAAAAATATTTGAGCTCAGTTACACTTGAATTTTACAGTACAGAA  
 2999 GCACTGGGATTTTATGTGCTTTTGTACCTTTTTCAGATCGGAATTAGTTTATGTTTAAAGCTTTAATGGTACTGATTTATGAAATGATGCATTAAGAGACAAGATATGTTGGGGTG

3119 GGGAGGAGTAAGGTGAATCACAATGCCTTCAATATGCTTTTGTACATCGCATGCTTTTATTAATAAAGGAGACTGAAAATTAATAAAAAAAAAAAAAAAAAA

**Figure 3.** Nucleotide and deduced amino acid sequence of chicken N-cadherin. The amino acid sequence determined by the direct sequencing of a purified tryptic fragment of N-cadherin (19) is shown with a broken line. The putative signal peptide and the putative membrane spanning region are shown with thick underlines. The potential N-linked glycosylation sites are indicated by arrowheads. The putative poly(A) addition signal is shown with a thin underline, and extension of  $\lambda$ N2 insert is indicated by arrows.

Interestingly, the morphology of L cell colonies was significantly altered by the expression of N-cadherin. In monolayer cultures, L cells barely form compact colonies and tend to migrate freely. However, the transformants expressing N-cadherin form clusters in which cells appeared to be tightly connected to each other. When these cell clusters were stained for N-cadherin, the antigens were concentrated at the boundary between cells, suggesting that they are directly involved in cell-cell connections (Fig. 5 c).

To examine whether the transformants have the functional N-cadherin, we performed a cell aggregation assay. L cells show a weak Ca<sup>2+</sup>-dependent aggregating activity when dissociated with 0.01% trypsin in the presence of 10 mM Ca<sup>2+</sup> and allowed to reaggregate (20). This activity, however, can be completely eliminated by treating the cells with the same concentration of trypsin in the presence of 1 mM Ca<sup>2+</sup> (TIC treatment). NL1 and NL2, dissociated by TIC treatment, showed strong aggregation in a Ca<sup>2+</sup>-dependent manner, and this activity was completely suppressed when the NCD-2 antibody specific to N-cadherin was added (Table I). These results indicate that N-cadherin derived from the cDNA possesses the cell-cell binding property by itself.

### Comparisons with Other Cadherins

We compared the amino acid sequence of N-cadherin with that of L-CAM (5), which is believed to be the chicken homologue of mouse E-cadherin, and with that of mouse E- and P-cadherin which was recently determined (15, 17). The results showed that there is a high degree of similarity in their primary structures throughout the entire length (Fig. 6).

Average similarity in amino acid sequences between the mature N-cadherin and L-CAM, both derived from the chicken but from different tissues, is 50%. This value is much smaller than the similarity between the mature E-cadherin and L-CAM (65%), the interspecies homologues. In other combinations, similarity between N-cadherin and E-cadherin, N-cadherin and P-cadherin, E-cadherin and P-cadherin, and L-CAM and P-cadherin is 49, 45, 58, and 57%, respectively. According to these results, N-cadherin has the most unique sequence among the four cadherins. N-cadherin is also distinguished from the others by insertions of amino acids at several points.

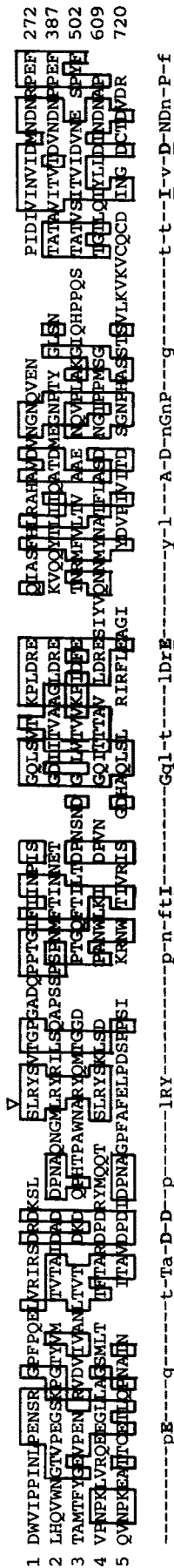
Figs. 6 and 7 compare the four cadherins. The putative extracellular domain of the mature proteins was divided into five segments having an about equal size. Among these segments, designated as EC1 to EC5, EC1 is most conserved, and EC5 is least conserved. EC5 is, however, characterized

**Table I. Ca<sup>2+</sup>-dependent Aggregation of Transformed and Untransformed L Cells**

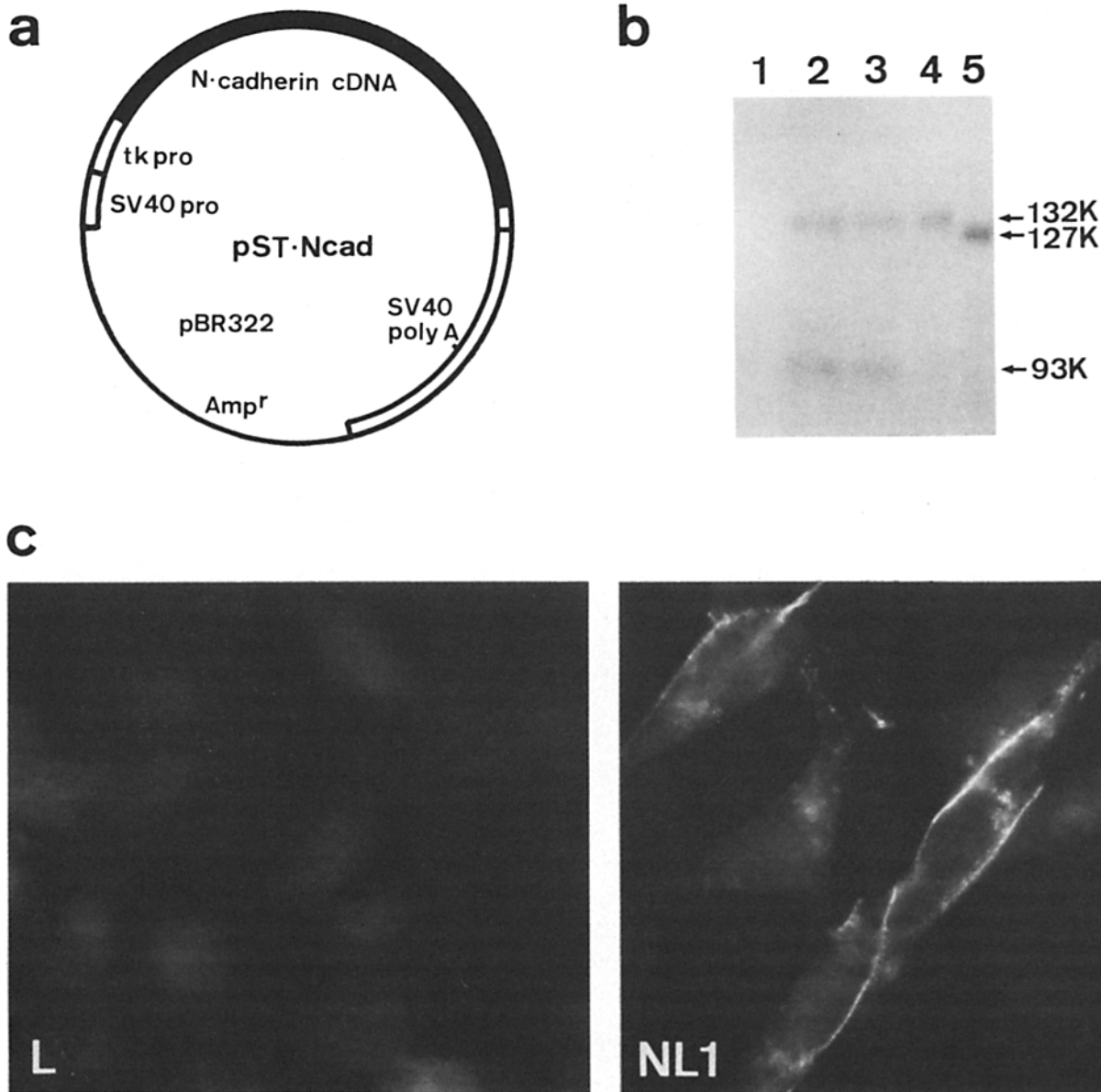
Cells	Aggregation index ( $N_{20}/N_0$ )		
	-Ca <sup>2+</sup>	+Ca <sup>2+</sup>	+Ca <sup>2+</sup> + NCD-2
L	1.0	1.0	1.0 (-)*
NL1	1.0	0.08	1.0 (100)*
NL2	1.0	0.19	0.92 (90)*

Percent inhibition was calculated as described (22).

\* Percent inhibition.



**Figure 4.** Internal repeats in the extracellular domain of N-cadherin. The amino acid sequence from position 165 to 720 was divided into five segments having an about equal size, and aligned to maximize the match among them. Residues identical in at least two segments are enclosed in boxes. Consensus repeats are shown below the sequences, where the underlined capital letters, the capital letters, and the small letters represent five repeats, four repeats, and three repeats, respectively. The computer analysis showed that the two stretches, each having 107 and 110 amino acids, starting from the leucine residues on segments 1 and 2 shown by a triangle, have the highest internal similarity (see the text and Fig. 7).



**Figure 5.** Expression of chicken N-cadherin in mouse L cells. (a) Construction of pST-Ncad used for transfection. (b) Western blot analysis of N-cadherin in untransformed L cells (lane 1), transformant NL1 cells (lane 2), transformant NL2 cells (lane 3), heart of a newly hatched chick (lane 4), and brain of a 5.5-d-old chicken embryo (lane 5). The Western blot analysis was performed as described previously (6). (c) Immunofluorescent detection of N-cadherin in untransformed L cells (left) and in NL1 cells (right). The monoclonal antibody NCD-2 and the rhodamine-conjugated anti-rat IgG were used for this analysis as described previously (9). Bar, 10  $\mu$ m.

**Figure 6.** Comparison of the predicted amino acid sequences of chicken N-cadherin (N), chicken L-CAM (L), mouse E-cadherin (E), and mouse P-cadherin (P). The entire amino acid sequence for each molecule was divided into the putative signal peptide (SIG), precursor region (PRE), extracellular region (EC), transmembrane region (TM), and cytoplasmic region (CP); the putative extracellular region was subdivided into five segments (EC1 to EC5). The sequences of the four molecules were then aligned to maximize the match; the five segments of the extracellular region were aligned in accordance with the internal repeats as shown in Fig. 4. Amino acid residues conserved among the four cadherin subclasses are shown in darkly shaded boxes, and those conserved among the three or two subclasses are shown in lightly shaded boxes. Cysteine residues conserved among all the subclasses are shown in dark boxes. Homologous repeats conserved among all the subclasses are marked with single or double dots; the double dots represent repeats between the adjacent segments and the single dot represents repeats between nonadjacent segments. Data for L-CAM, E-cadherin, and P-cadherin were taken from Gallin et al. (5), Nagafuchi et al. (15), and Nose et al. (17), respectively.

SIG N MCR IAGTTPRIPPLALMLLAALQOAPI  
L E MGARCRSFSALLLLQVSWMLCQLEP  
P MELLSGPHAFLLLLLQVCWLSRVYSEP

PRE N KATCEDMLCKMGFPEDY HSAVVRSRVHGQPELVNRFQSCDENRKYIYSSSEPEDEFRVGEDGVVYVAERSFQLSAETEFVVSARDKETEQEEQMKVLT PEPAFTGASEKDKKIEDIIFPQWQKDSSSHKEQKR  
L ---EFRDVAAGRELGKRVSAACS GRPVAVYPTDTRFKVNGDGVVSTKPLTYLGRKISFTIYAQDAMGR HSARVTV GRHRRRHHHHHHHLDQDTPAVLTFPKHDPGFLRQKR  
E ESCSPGFSSEV YTFPVERHLERGHVLRVFECT GRPRTAFFSDESRKVAQDGTITVXRHLKHLKLETSFLVLRARDSSHRE LSTKVTLKSMGHRRHHRRHPASENPPELLMFPVYVPG LREQKR  
P YRAGFIGEAGVTEVEGTDLEPSQVIGKVALAQG MHHADNGDIITMLTRGTGQGGKDHAMSPPTLRIRRRKR

EC1 N DWYIPFANLPSNR GPFQFELVRAIRSDRDKSL SLKYSVGPAGADQPFQIFMINPIS  
L DWYIPFASCLNHR GPYEMRLVQIKSNKOKES KVVYSINGQAGADSPVGI FIERET  
E DWYIPFASCPENK GEFKFNLVQIKSNRDKET KVFYSINGQAGADKPPVGV FIERET  
P EWTMPPFVPEMCK GPFQFQRLNQLKSNKDRGT KIFYSINGQAGADSPVGV FIERES

EC2 N LHQVWNGTVPEGSKPCTIYM TWTADAD DPAQNGMLRKRILSQAESSPSPNMTINNET  
L IKEVFVGYIEENAKPCTISM TYNATDADDVNTQNGIVSYSVSQPPRPHQMFETIDPAK  
E TQVFEGSVAEGAVPCTISM KVSATDADDVNTYNAALATYIVSQDELPKMMFTVNRDT  
P TQDTFRGSLVLEGVMPCTISM QVTAIDEDDVAVTVNGVAVSIHSQEKERHDLMEFIHKST

EC3 N NPTMVFYGEVPE RVDVIVANLIVT DKD QPHTPAWNRKRYQMTGGD PTGQFTRILITPNSND G LVTVVVKPIDFE  
L NPTMVFYGEVPEEN PCGTEVARLIVT DQD AFGSPAQAVYHIKSGN LDGAKSIIITPSTNN GILKTA KGLDYE  
E NPSTYQGVPEEN EVNARIATKVT DDD AENTPAKAVYTVV ND PQQFVVVHTPTVND GILKTA KGLDFE  
P EPQKYEAWVPEEN EVGHEVQRIVT DLD VENWPAWARYHIHVGDD DGDHFTITTHPEINQ GVLITK KGLDFE

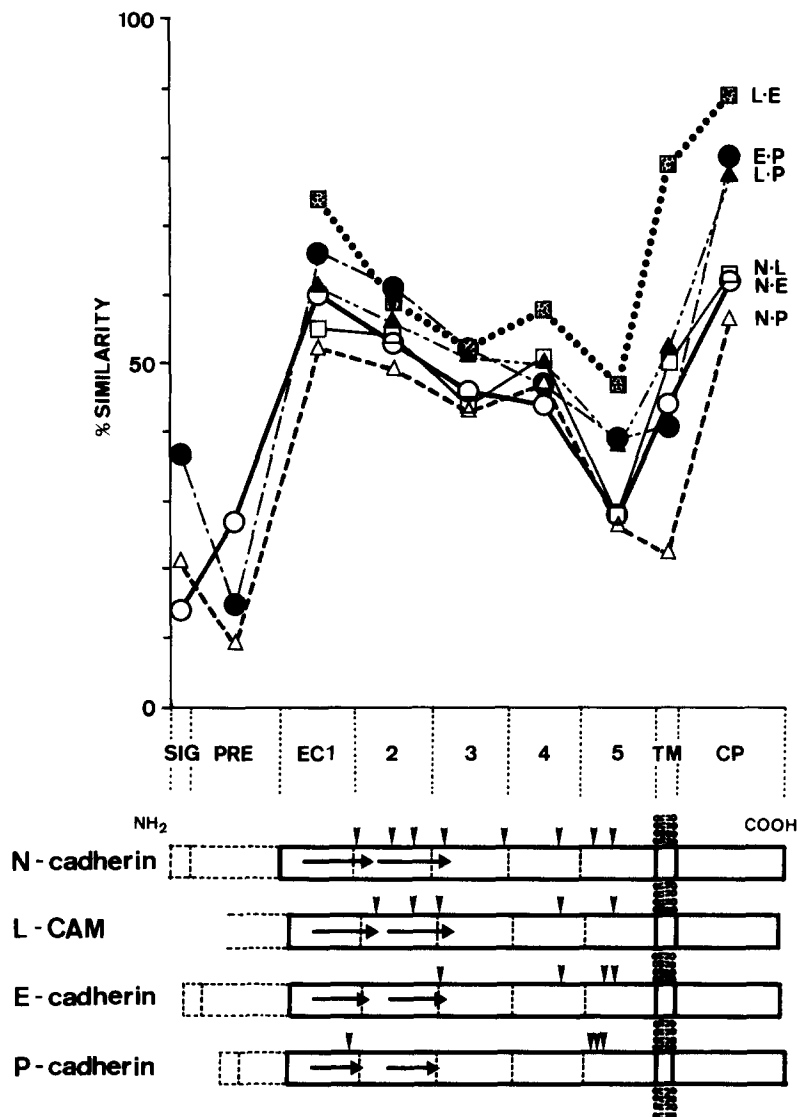
EC4 N VNPDKLVQEBGLLAGSMLT TFRDPRDMQOT SLRASKLSR PANWIKI DEVN GQITTVAV LDRESI YVONWNAIFLAD NGTTPMSG  
L VPTPKRYGVPEDELVPVQOVT SVTAQDPDRMRQ KIVWRMGSD PAGWLYI HPEM G IVTATQPLDRESV HAINSTWKAIIAVD NGTPTDTG  
E MPARRRVEVPEDFGVQOET SVTAREPTMDQ KITVRIWRD TANWLEI NPET GAIPTKAE MDREDAEHVKNSTVALLIATD DGEFIATG  
P VPSPKIVIAQEGISIGELVC IYTAQDPDKQ KISVITISRD PANWIAV DEPS GOI TAAAGILDRDEQFVKNNVYEVAVLAVD SGNPPTTG

EC5 N QVNEKEATTCETLQPNAIN ITAVDPDIDBNAGPFAFEPDPSPSI KRNW TIVRIS GDHAQLSL RIRFLEAGI YDVPVIVTD SGNPHASSTVLRKVKVQOD ING DQTDVDR  
L TPERSEFICSR QPEKQI LSVDKDLPPHTYPFKAALDEHGSS NNW TVEIRG QDELAMGL KKELEPGE YNIFVKLTD SQK AQVTQV KAQVCSGGTAKN GERRSY  
E IPEPRNQFCQR NPQPHI ITWLDPLPNTSPFTAETHGAS VNW TIEYNDAAQESLILQP RKDLEIGE YKHLKLAB NQNKD QVT TLDVHVCSGGTVANN CMK  
P IPEPRQIICNQ SRVQV LNIITDKDLSNPSFPQAQIHTHSD IYW MAEVSE KGDVVALSL KFKLKQDT YDLHLSLSD HGNRE QLT MIRATVCSGGVFNDCPRPWK

TM N IVNAGLGTGAIIALLCIIILLLVLMFVVMW  
L IVG GLGVRAILGILGGILALMLLMLLLFLA  
E AGIVAAQLQVPAALGILGGILALMLLMLLLFL  
P G GFILP ILGAV LALMLLALMLLLLV

CP N KRRDKERQAKQLID PEDDVRDNILKYDEEGGEEQDYDLSQLQ QPDTVEPAIKPVGIRGLDRP IHAEBQVYVRSNAP  
L RRRKVKKEP LLP PEDDVRDNVNYDEEGGEEQDYDLSQLHGLDAR PEVI RNDVAPPLMAA PQRPRPA  
E RRRKVKKEP LLP PDDETRDNVYDEEGGEEQDYDLSQLHGLDAR PEVI RNDVAPPLMSV PQRPRPA  
P RRRKVKKEP LLP PEDDTRDNVYDEEGGEEQDYDITQLHRLGLAR PEVVL RNDVYVTFIPTPMVPRPA

N HFGIDGFTNEGKRAANDPTAPPYDLSLVFDYEGSGSTAGLSLNSNSGGEQDYDLYLNDWGRPRFKKLADWYGGG DD  
L NPDEIGNFIDENLKAADTPTAPPYDLSLVFDYEGSGSEATLSLNSNSADQDYDLYLWGNRFRKKLALYGGGEDD  
E NPDEIGNFIDENLKAADPTAPPYDLSLVFDYEGSGSEAA LSLNSNSADQDYDLYLWGNRFRKKLADWYGGGEDD  
P NPDEIGNFIDENLKAANTPTAPPYDLSLVFDYEGSGSDRALSLSLTTSASDQDYDLYLWGNRFRKKLADWYGGGEDD



**Figure 7.** Similarities in amino acid sequences among four cadherin subclasses in different regions of the peptides. Sequences from these molecules are divided into nine regions as defined in Fig. 6. Percent similarity was calculated for each region for each pair of cadherin subclasses. A schematic representation of the overall structure of the four cadherin subclasses is also shown. The major internal repeats detected by the computer search using the IDEAS program (11) are shown with arrows. Potential N-linked glycosylation sites are shown with arrowheads. *L*, L-CAM; *E*, E-cadherin; *P*, P-cadherin; *N*, N-cadherin. Data for L-CAM, E-cadherin, and P-cadherin were taken from the same papers cited in the legend to Fig. 6.

by conservation of four cysteines in all four cadherins. The internally repeated sequences containing negatively charged amino acids, such as D-X-D, L-D-R-E, and D-X-N-D/E, are also conserved (Fig. 6).

The most conserved regions for all four cadherins are in the cytoplasmic domain (Figs. 6 and 7). N-cadherin, however, again had unique sequences when compared with the other cadherins. The precursor sequences preceding the putative NH<sub>2</sub> terminus were rather unique for each cadherin subclass, although some similarity was found, particularly in the flanking region of the putative NH<sub>2</sub> terminus.

### Discussion

The finding that the amino acid sequences, as well as the topologies, are similar between N-cadherin and three other molecules, L-CAM, E-cadherin, and P-cadherin, confirms our hypothesis that these molecules belong to the same gene family. This family seems to be an entirely new gene family, since no similar sequences in other proteins were found in databases.

Comparisons of sequences among cadherin subclasses allow us to assess the functional significance of different regions of each peptide. Similarities in sequence among the subclasses might reflect their common properties, whereas differences may reflect their specificities.

The most conserved property of cadherins is that of Ca<sup>2+</sup> binding. We could not detect the known consensus sequence for the Ca<sup>2+</sup>-binding site, called the E-F hand, in cadherin sequences. However, characteristic sequences with negatively charged amino acids were conserved in the form of internal repeats. These sequences may be candidates for the Ca<sup>2+</sup>-binding sites. The conservation of cysteine residues in the proximal region of the extracellular domain must also play an important role for maintaining the general structure of cadherins.

The highest similarity among the four cadherins was found in the putative cytoplasmic domain, suggesting that this region is important in cadherin function. Our recent studies suggested that cadherins are associated with cytoskeletal components (9). Therefore, the highly conserved cytoplasmic sequences may play a role in the interaction with such cytoplasmic elements.



It is difficult to determine the sequences responsible for cell-cell binding specificities of cadherin subclasses only by comparing their sequences. The following, however, is an important point: The monoclonal antibody NCD-2 recognized a fusion protein translated from the cDNA probe ( $\lambda$ NI) encoding the putative NH<sub>2</sub>-terminal region. It is known that this antibody specifically blocks the N-cadherin-mediated cell adhesion. The NH<sub>2</sub>-terminal region, therefore, might contain sequences responsible for the N-cadherin-specific cell-cell binding sites. The same was observed with a monoclonal antibody PCD-1 specifically blocking the P-cadherin-mediated adhesion; this antibody recognized the region around the NH<sub>2</sub> terminus (16). On the other hand, the NH<sub>2</sub>-terminal region shows the highest sequence similarity among cadherin subclasses, suggesting that this region is important for the general function of cadherins. This apparent contradiction suggests that the sequences responsible for the common function and the subclass specificities may not be clearly segregated from each other.

It should be interesting to compare the sequences of cadherins with respect to the subclass specificities versus the animal species specificity. Similarity between N-cadherin and L-CAM, both derived from the chicken, is 50% and that between E- and P-cadherin, both derived from mouse, 58%. Although L-CAM is thought to be the chicken homologue of E-cadherin, similarity between these molecules is only 65%. This unexpectedly low similarity between L-CAM and E-cadherin suggests that, although these molecules might have originally been more closely related to each other than to other cadherin subclasses, they have diverged to a large extent in different species. This divergence might have produced species specificity of cadherin molecules; there were observations that mouse cells segregate from chicken cells derived from homologous tissues when mixed artificially (1, 2).

The present success in the ectopic expression of N-cadherin cDNA provides a powerful experimental system to investigate the molecular mechanism of cadherin-mediated cell adhesion. We have also recently succeeded in expressing E-cadherin cDNA in L cells (15). Using these systems, it is now possible to analyze the relationships between the structure and function of cadherins by genetically dissecting these molecules. These approaches should also be very useful for investigating the molecular basis of specificity of the cadherin subclasses in future studies.

We wish to thank Dr. Kunio Yasuda for helpful technical advice, and Drs. Kazuto Katoh and Hisato Kondoh for providing a plasmid vector. We also thank Mr. Masaya Hosoda for his help in preparing antibodies.

This work was supported by the Special Coordination Fund of the Science and Technology Agency of the Japanese Government and also by the research grants from the Ministry of Education, Science and Culture of Japan. K. Hatta is a recipient of a Fellowship of the Japan Society for the Promotion of Science for Japanese Junior Scientists.

Received for publication 21 September 1987, and in revised form 15 December 1987.

## References

- Burdick, M. L. 1970. Cell sorting out according to species in aggregates containing mouse and chick embryonic limb mesoblast cells. *J. Exp. Zool.* 175:357-368.
- Burdick, M. L., and M. S. Steinberg. 1969. Embryonic cell adhesiveness: do species differences exist among warmblooded vertebrates? *Proc. Natl. Acad. Sci. USA.* 63:1169-1173.
- Cunningham, B. A., Y. Leutzinger, W. J. Gallin, B. C. Sorkin, and G. M. Edelman. 1984. Linear organization of the liver cell adhesion molecule L-CAM. *Proc. Natl. Acad. Sci. USA.* 81:5787-5791.
- Duband, J.-L., S. Dufour, K. Hatta, M. Takeichi, G. M. Edelman, and J. P. Thiery. 1987. Adhesion molecules during somitogenesis in the avian embryo. *J. Cell Biol.* 104:1361-1374.
- Gallin, W. J., B. C. Sorkin, G. M. Edelman, and B. A. Cunningham. 1987. Sequence analysis of a cDNA clone encoding the liver cell adhesion molecule, L-CAM. *Proc. Natl. Acad. Sci. USA.* 84:2808-2812.
- Hatta, K., and M. Takeichi. 1986. Expression of N-cadherin molecules associated with early morphogenetic events in chick development. *Nature (Lond.)* 320:447-449.
- Hatta, K., T. S. Okada, and M. Takeichi. 1985. A monoclonal antibody disrupting calcium-dependent cell-cell adhesion of brain tissues: Possible role of its target antigen in animal pattern formation. *Proc. Natl. Acad. Sci. USA.* 82:2789-2793.
- Hatta, K., S. Takagi, H. Fujisawa, and M. Takeichi. 1987. Spatial and temporal expression pattern of N-cadherin cell adhesion molecules correlated with morphogenetic processes of chicken embryos. *Dev. Biol.* 120: 215-227.
- Hirano, S., A. Nose, K. Hatta, A. Kawakami, and M. Takeichi. 1987. Calcium-dependent cell-cell adhesion molecules (cadherins): subclass-specificities and possible involvement of actin bundles. *J. Cell Biol.* 105:2501-2510.
- Huynh, T. V., R. A. Young, and R. W. Davis. 1985. Constructing and screening cDNA libraries in  $\lambda$ gt10 and  $\lambda$ gt11. In *DNA cloning: A Practical Approach*. D. M. Glover, editor. IRL Press, Oxford. 49-78.
- Kanehisa, M., and W. B. Goad. 1982. Pattern recognition in nucleic acid sequences. II. An efficient method for finding locally stable secondary structures. *Nucleic Acids Res.* 10:265-278.
- Katoh, K., Y. Takahashi, S. Hayashi, and H. Kondoh. 1988. Improved mammalian vectors for high expression of G418 resistance. *Cell Struct. Funct.* In press.
- Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157:105-132.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular cloning. A laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. 1-545.
- Nagafuchi, A., Y. Shirayoshi, K. Okazaki, K. Yasuda, and M. Takeichi. Transformation of cell adhesion properties by exogenously introduced E-cadherin cDNA. *Nature (Lond.)* 329:341-343.
- Nose, A., and M. Takeichi. 1986. A novel cadherin cell adhesion molecule: its expression pattern associated with implantation and organogenesis of mouse embryos. *J. Cell Biol.* 103:2649-2658.
- Nose, A., A. Nagafuchi, and M. Takeichi. Isolation of placental cadherin cDNA: identification of a novel gene family of cell-cell adhesion molecules. *EMBO (Eur. Mol. Biol. Organ.) J.* 6:3655-3661.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463-5467.
- Shirayoshi, Y., K. Hatta, M. Hosoda, S. Tsunasawa, F. Sakiyama, and M. Takeichi. 1986. Cadherin cell adhesion molecules with distinct binding specificities share a common structure. *EMBO (Eur. Mol. Biol. Organ.) J.* 5:2485-2488.
- Takeichi, M. 1987. Cadherins: a molecular family essential for selective cell-cell adhesion and animal morphogenesis. *Trends Genet.* 3:213-217.
- Takeichi, M., T. Atsumi, C. Yoshida, K. Uno, and T. S. Okada. 1981. Selective adhesion of embryonal carcinoma cells and differentiated cells by Ca<sup>2+</sup>-dependent sites. *Dev. Biol.* 87:340-350.
- Urushihara, H., H. S. Ozaki, and M. Takeichi. 1979. Immunological detection of cell surface components related with aggregation of chinese hamster and chick embryonic cells. *Dev. Biol.* 70:206-216.
- Xie, H-X. 1984. Differences in the efficiency and stability of gene expression after transfection and nuclear injection: a study with a chick  $\delta$ -crystallin gene. *Cell Struct. Funct.* 8:315-325.
- Yoshida-Noro, C., N. Suzuki, and M. Takeichi. 1984. Molecular nature of the calcium-dependent cell-cell adhesion system in mouse teratocarcinoma and embryonic cells studied with a monoclonal antibody. *Dev. Biol.* 101:19-27.