The recent explosion in genome sequencing has revealed the great diversity of the cadherin superfamily. Within the superfamily, protocadherins, which are expressed mainly in the nervous system, constitute the largest subgroup. Nevertheless, the structures of only the classical cadherins are known. Thus, to broaden our understanding of the adhesion repertoire of the cadherin superfamily, we determined the structure of the N-terminal first extracellular cadherin (EC) domain, which is conserved among the classical (type I), type II, and desmosomal cadherins (5–13). Most of the previous functional and structural analyses have been of the classical cadherins. These studies revealed homophilic adhesive binding interfaces localized primarily within the N-terminal first extracellular cadherin (EC) domain, which is conserved among the classical (type I), type II, and desmosomal cadherins (5–13). However, the classical cadherins account for only a fraction of the cadherin superfamily, which has a multitude of diverse members. Protocadherins are now known to constitute the largest subgroup within the cadherin superfamily (see Fig. 1A) (4, 9, 11, 14). Most protocadherins have a divergent cytoplasmic domain and six or seven EC domains, with low sequence similarities to the EC domains of the classical cadherin group (15). Here, we have focused on one of the major cluster-type protocadherins, the cadherin-related neuronal receptor/protocadherin-α (CNR/Pcdhα) family. The CNR/Pcdhα genome is organized into an unusual gene cluster that is similar to the organization of the immunoglobulin and T-cell receptor genes.

In recent years, the cadherins have emerged as an important superfamily, and their structures and biological functions are proving to be complex. Originally thought of as calcium-dependent cell adhesion molecules, the cadherin superfamily molecules are now known to be involved in many biological processes, including cell recognition, cell signaling, cell communication during embryogenesis, and the formation of neural circuits in the central nervous system (1–3). The cadherin superfamily can be divided into several subgroups (4): the classical (type I) and closely related type II cadherins, the desmosomal cadherins, and the protocadherins (see Fig. 1A). Most of the previous functional and structural analyses have been of the classical cadherins. These studies revealed homophilic adhesive binding interfaces localized primarily within the N-terminal first extracellular cadherin (EC) domain, which is conserved among the classical (type I), type II, and desmosomal cadherins (5–13).
(see Fig. 1B) (16). The mouse CNR/Pcdha gene cluster is composed of 14 variable-region exons and a set of three constant-region exons (17, 18). Mature CNR/Pcdha mRNAs are generated from one of these variable-region exons and the three constant-region exons, and their differential and combinatorial expression is observed at the individual neuron level (19–21).

Among the six tandemly repeated EC domains of the CNR/Pcdha protein, the N-terminal EC1 domain shows several unique features. First, the sequence of the EC1 domain is well conserved among mouse CNR/Pcdha family members (Fig. 1B) (19). Second, the EC1 domain has a sequence containing the RGD motif, which is known to function in protein–protein interactions. Recently, the RGD motif of the CNR/Pcdha EC1 domain was shown to be involved in the adhesion activity of the CNR/Pcdha EC1 domain in HEK293T cells, which occurs via β1 integrin (22). Third, the CNR/Pcdha EC1 domain lacks Trp2, which is highly conserved among the classical (type I), type II, and desmosomal cadherins and is critical for their homophilic binding activity (11). In fact, CNR/Pcdha appears to possess no homophilic binding activity (22).

These characteristics of the CNR/Pcdha EC1 domain raise the possibility that the EC1 domains of the cadherin superfamily, which are thought to be essential for adhesion, may be structurally and functionally rather divergent. However, this issue has been difficult to investigate because no structures of cadherin superfamily EC1 domains, except those of classical cadherins, have been determined. Furthermore, current knowledge about the functions of CNR/Pcdha is still meager compared with what is known about the classical cadherins. Here, to better understand the adhesion repertoire if the cadherin superfamily and to help elucidate the unknown functions of the CNR/Pcdha family proteins, we determined the solution structure of the CNR/Pcdha N-terminal EC1 domain by NMR. We further investigated the function of the domain on the basis of its determined structure.

**EXPERIMENTAL PROCEDURES**

**NMR Sample Preparation**—The DNA encoding the EC1 domain of mouse CNR/Pcdha (Glu1–Asn103) was amplified by PCR from mouse CNR/Pcdha cDNA and subcloned into the pET-11d vector (Novagen). The recombinant protein was expressed from the resulting plasmid, which contained an additional methionine at the N terminus. The CNR/Pcdha EC1 domain was labeled uniformly with stable 15N and 13C isotopes by PCR from mouse CNR/Pcdha cDNA and subcloned into the pET-11d vector with a NcoI restriction enzyme site. The DNA encoding the EC1 domain was digested with NcoI and cloned into the pET-11d vector. The recombinant protein was expressed from the resulting plasmid, which contained an additional methionine at the N terminus.

The recombinant protein was expressed from the resulting plasmid, which contained an additional methionine at the N terminus. The CNR/Pcdha EC1 domain was labeled uniformly with stable 15N and 13C isotopes by PCR from mouse CNR/Pcdha cDNA and subcloned into the pET-11d vector with a NcoI restriction enzyme site. The DNA encoding the EC1 domain was digested with NcoI and cloned into the pET-11d vector. The recombinant protein was expressed from the resulting plasmid, which contained an additional methionine at the N terminus.

**NMR Sample Preparation**—The DNA encoding the EC1 domain of mouse CNR/Pcdha (Glu1–Asn103) was amplified by PCR from mouse CNR/Pcdha cDNA and subcloned into the pET-11d vector (Novagen). The recombinant protein was expressed from the resulting plasmid, which contained an additional methionine at the N terminus. The CNR/Pcdha EC1 domain was labeled uniformly with stable 15N and 13C isotopes by PCR from mouse CNR/Pcdha cDNA and subcloned into the pET-11d vector with a NcoI restriction enzyme site. The DNA encoding the EC1 domain was digested with NcoI and cloned into the pET-11d vector. The recombinant protein was expressed from the resulting plasmid, which contained an additional methionine at the N terminus.

**Resonance Assignments**—To assign the 1H, 15N, and 13C resonances, a series of two- and three-dimensional experiments were performed. These were 15N-1H HSQC-wg, 13C-1H HSQC-se, and 1H-1H TOCSY with a mixing time of 57.6 ms; 15N-edited TOCSY-se with a mixing time of 69.0 ms; and HNCO-wg, HN(CA)CO-se, CBCA(CO)NH-se, HN(CA)CO-se, CBCA(CO)NH-se, C(CO)NH-se, H(CO)NH-se, and HCCH TOCSY with mixing times of 20.2 ms each, where “wg” refers to the WATERGATE and water flip-back method and “se” refers to the sensitivity enhancement and gradient echo method (24). In the NMR experiments in which amide protons were detected directly, the spectral width of the 1H dimension was set to 14.0 ppm. The 1H carrier was set at the frequency of the residual water resonance (4.7 ppm), and the 15N carrier was set at 121.0 ppm. The acquisition times were 20 and 71 ms for the 15N and 1H dimensions, respectively, independent of the scales of the static magnetic fields applied. The other indirect dimensions were acquired using the TIPPI-States method (24).

All data were processed with the program NMRPipe (25). The peaks were analyzed with the program Sparky (developed by T. D. Goddard and D. G. Kneller, University of California, San Francisco). The assignment results and accompanying details have been described elsewhere (26) and deposited in the BioMagResBank under accession number 6405. A number of resonances were missing in the 1H-15N HSQC spectra (Glyγ, Asnγ, Serγ, Gluγ, Thrγ, Lysγ, Glyγ, Gluγ, Argγ, Alaγ, Cysγ, and Aspγ), probably because of rapid amide proton exchanges with the solvent at pH 8.0 and conformational exchanges (26). Samples prepared at a low pH (6.0–7.0) immediately formed precipitates, which prevented us from detecting amide resonances for these residues. Relaxation analyses for amide 15N spins were performed as described (27, 28).

**Structural Restraints and Calculations**—To obtain interproton distance restraints, two-dimensional NOESY, three-dimensional 15N-edited NOESY, and 13C-edited NOESY with mixing times of 100 ms each were acquired. The NOE connectivities derived from strong, medium, and weak cross-peaks were categorized and assumed to correspond to the upper limits for the interproton distances of 3.0, 4.0, and 5.0 Å, respec-
Structure of the EC1 Domain of CNR/Protocadherin-α

A

Type I cadherins
Desmosomal cadherins
Type II cadherins
Protocadherins
CNR/Pcdhα family
Pcdhγ family
Pcdhβ family

B

DNA
mRNA splicing
Full-length protein
EC1 domain protein

Protocadherin Family Gene Clusters
TABLE 1
NMR structural statistics for the CNR/Pcdhα EC1 domain

These statistics comprise the ensemble of the final 30 simulated annealing structures of 100 calculated with CYANA. Values are means ± S.D. r.m.s., root mean square deviation.

| Total no. of distance restraints | 1100 |
|----------------------------------|------|
| Intraresidual                   | 463  |
| Sequential | i − j | 211 |
| Medium range | i − j | ≤ 4 | 93 |
| Long range | i − j | > 4 | 333 |
| Hydrogen bonds                  | 40   |
| Disulfide bond                   | 1    |
| No. of dihedral angle restraints | 185  |
| Distance restraint violations >0.5 Å | 0 |
| Dihedral angle restraint violations >5° | 0 |
| van der Waals distance violations >0.5 Å | 0 |
| r.m.s. deviations from experimental restraints |
| Distance (Å)                     | 0.0170 ± 0.0012 |
| Angle                            | 0.4107 ± 0.0579 |
| PROCHECK Ramachandran plot statistics |
| (residues 1–103; %)              |      |
| Residues in most favored regions | 77.8 |
| Residues in additional allowed regions | 21.1 |
| Residues in generously allowed regions | 1.1 |
| Residues in disallowed regions   | 0.0  |
| r.m.s. deviations from mean coordinate positions |
| Full-length (residues 1–103, Å)  |      |
| Backbone heavy atoms             | 0.631 ± 0.128 |
| All heavy atoms                  | 1.185 ± 0.116 |
| Secondary structure regions (residues 4–10, 18–20, 38–41, 49–51, 58–60, 77–84, and 89–98, Å) |
| Backbone heavy atoms             | 0.251 ± 0.067 |
| All heavy atoms                  | 0.814 ± 0.102 |

Structure of the EC1 Domain of CNR/Protocadherin-α

Preparation of Fc Fusion Proteins for Functional Assays—The cDNAs of full-length mouse N-cadherin and CNR/Pcdhα were amplified by reverse transcription-PCR and subcloned into the pBS vector. The CNR/Pcdhα EC1/N-cadherin EC2–5 (CNRREC1/Ncad) and N-cadherin EC1/CNR/Pcdhα EC2–5 (NcadEC1/CNR) chimeric constructs were generated by PCR ligation from the full-length N-cadherin and CNR/Pcdhα constructs. The N-cadherin EC1–5–Fc (Ncad-Fc), N-cadherin (signal sequence + prodomain)-Fc (Fc), CNR/Pcdhα EC1–6–Fc (CNR/Pcdhα-Fc), CNR/Pcdhα EC1–FC, CNREC1/Ncad-Fc, and NcadEC1/CNR-Fc constructs were derived from the full-length N-cadherin, CNR/Pcdhα, and CNREC1/Ncad chimeric constructs, described above, by PCR to amplify the entire extracellular domains or smaller regions and subsequently subcloned into the pIB/Fc vector (a gift from Dr. Zipurksy, UCLA) via NotI/SpeI sites. These constructs were subcloned into pcDNA3.1 via NotI/XbaI sites for mammalian cell expression. The chimeric Fc proteins used for the aggregation and cell adhesion assays were purified from HEK293T cells as follows. The cells were transfected using Lipofectamine 2000 (Invitrogen) in Dulbecco’s modified Eagle’s medium and 10% ultra low protein fetal bovine serum. After 1 week in culture, the cleared medium was incubated overnight with nProtein A-Sepharose 4 Fast Flow beads (Amersham Biosciences) to capture the Fc-tagged proteins. The Protein A beads were washed with phosphate-buffered saline, applied to an open column, eluted with 100 mM glycine (pH 2.8), and subsequently neutralized with 1 M Tris (pH 9.5).

Bead Aggregation Assay—Bead aggregation was assayed by a modification of a method described previously (35). In brief, mouse anti-human IgG Fc (Chemicon International) was passively absorbed to red fluorescent microspheres (0.39-μm beads; Duke Scientific Corp.). The IgG-coupled beads were washed with 20 mM HEPES (pH 7.2) containing 100 mM NaCl and 5% fetal bovine serum and then with 20 mM HEPES (pH 7.2) containing 100 mM NaCl and 0.1% bovine serum albumin (BSA) (Sigma). One microliter of IgG-coupled beads was coated with purified Fc-tagged protein (250 nm in a volume of 100 μl) in 20 mM HEPES (pH 7.2) containing 100 mM NaCl, 1 mM CaCl₂, and 0.1% BSA overnight at 4 °C with gentle agitation. Following capture of the chimeric Fc proteins, the beads were sonicated for 5 min in an ice-bath cup horn (Misonix) in polystyrene tubes. The beads were then incubated at 25 °C with shaking (1400 rpm). At various time points, 3-μl aliquots of the beads were transferred into 600 μl of 20 mM HEPES (pH 7.2) containing 100 mM NaCl, 1 mM CaCl₂, and 0.1% BSA, and 20,000 events were analyzed by FACS (EPICS ALTRA, Beckman Coulter). Following the FACS analysis, 4 μl of beads were spotted onto a glass microslide for visualization. To assess the levels of Fc-tagged proteins on the beads, the beads were pelleted after tively (26). The distance restraints for the hydrogen bonds were applied for the amides that were judged to form a β strand on the basis of the characteristic NOE patterns and the φ and ψ dihedral angles predicted using the software TALOS (29), using 2.8–3.3 Å for nitrogen-oxygen pairs and 1.8–2.3 Å for hydrogen-oxygen pairs, after the root mean square deviations for the overlaid backbone nuclei of calculated structures reached 1.0 Å. The backbone torsion angles were predicted using TALOS (29), CSP vector, the TALOS/CSP vector and subcloned into the pBS vector. The CNR/Pcdhα EC1/N-cadherin EC2–5 (CNRREC1/Ncad) and N-cadherin EC1/CNR/Pcdhα EC2–5 (NcadEC1/CNR) chimeric constructs were generated by PCR ligation from the full-length N-cadherin and CNR/Pcdhα constructs. The N-cadherin EC1–5–Fc (Ncad-Fc), N-cadherin (signal sequence + prodomain)-Fc (Fc), CNR/Pcdhα EC1–6–Fc (CNR/Pcdhα-Fc), CNR/Pcdhα EC1–FC, CNREC1/Ncad-Fc, and NcadEC1/CNR-Fc constructs were derived from the full-length N-cadherin, CNR/Pcdhα, and CNREC1/Ncad chimeric constructs, described above, by PCR to amplify the entire extracellular domains or smaller regions and subsequently subcloned into the pIB/Fc vector (a gift from Dr. Zipurksy, UCLA) via NotI/SpeI sites. These constructs were subcloned into pcDNA3.1 via NotI/XbaI sites for mammalian cell expression. The chimeric Fc proteins used for the aggregation and cell adhesion assays were purified from HEK293T cells as follows. The cells were transfected using Lipofectamine 2000 (Invitrogen) in Dulbecco’s modified Eagle’s medium and 10% ultra low protein fetal bovine serum. After 1 week in culture, the cleared medium was incubated overnight with nProtein A-Sepharose 4 Fast Flow beads (Amersham Biosciences) to capture the Fc-tagged proteins. The Protein A beads were washed with phosphate-buffered saline, applied to an open column, eluted with 100 mM glycine (pH 2.8), and subsequently neutralized with 1 M Tris (pH 9.5).

Bead Aggregation Assay—Bead aggregation was assayed by a modification of a method described previously (35). In brief, mouse anti-human IgG Fc (Chemicon International) was passively absorbed to red fluorescent microspheres (0.39-μm beads; Duke Scientific Corp.). The IgG-coupled beads were washed with 20 mM HEPES (pH 7.2) containing 100 mM NaCl and 5% fetal bovine serum and then with 20 mM HEPES (pH 7.2) containing 100 mM NaCl and 0.1% bovine serum albumin (BSA) (Sigma). One microliter of IgG-coupled beads was coated with purified Fc-tagged protein (250 nm in a volume of 100 μl) in 20 mM HEPES (pH 7.2) containing 100 mM NaCl, 1 mM CaCl₂, and 0.1% BSA overnight at 4 °C with gentle agitation. Following capture of the chimeric Fc proteins, the beads were sonicated for 5 min in an ice-bath cup horn (Misonix) in polystyrene tubes. The beads were then incubated at 25 °C with shaking (1400 rpm). At various time points, 3-μl aliquots of the beads were transferred into 600 μl of 20 mM HEPES (pH 7.2) containing 100 mM NaCl, 1 mM CaCl₂, and 0.1% BSA, and 20,000 events were analyzed by FACS (EPICS ALTRA, Beckman Coulter). Following the FACS analysis, 4 μl of beads were spotted onto a glass microslide for visualization. To assess the levels of Fc-tagged proteins on the beads, the beads were pelleted after
the FACS analysis; washed twice with 20 mM HEPES (pH 7.2) containing 100 mM NaCl, and 0.1% BSA; boiled in SDS sample buffer; and analyzed by SDS-PAGE, with subsequent immunoblotting using anti-Fc antibody. All of the protein-coated beads used in the assay were found to contain equivalent levels of purified proteins (see Fig. 5C).

Cell Adhesion Assay—Microtiter plates (96-well; Nunc) were coated with each protein (200 nM) dissolved in 20 mM HEPES (pH 7.2), 100 mM NaCl, and Hanks’ balanced saline solution at 40 μl/well for 1 h at 37 °C in a humidified CO2 incubator. The wells were subsequently washed twice with Hanks’ balanced saline solution and then blocked with 1% BSA and phosphate-buffered saline at 37 °C for 2 h. HEK293T cells were washed twice with phosphate-buffered saline and treated with phosphate-buffered saline containing 2 mM EDTA for 5 min at 37 °C. The cells were collected and washed with 20 mM HEPES (pH 7.2) containing 137 mM NaCl, 3 mM KCl, and 0.1% BSA. The pellets were suspended again at 1 × 10^6 cells/ml. Aliquots of the cell suspension were placed on the test...
dishes with 1 mM CaCl$_2$ and 1 mM MgCl$_2$. To activate $\beta_1$ integrin, antibody TS2/16 was added at 10 $\mu$g/ml for 15 min before plating.

The plated cells were incubated for 20 min and photographed by a camera fitted onto a microscope (Olympus IX70/DP50 system). The attached cells were counted in two independent fields per well.

RESULTS AND DISCUSSION

Overall Structure of the EC1 Domain of CNR/Pcdh$\alpha$ and Comparison with Those of Other Cadherins—The solution structure of the CNR/Pcdh$\alpha_4$ EC1 domain was determined by multidimensional double and triple resonance NMR spectroscopy. The stability of the protein could be maintained only at a low protein concentration of 0.1 mM and a high pH of 8.0, conditions that are generally not suitable for NMR measurements. We overcame this problem by using a high magnetic field NMR spectrometer (800 MHz) with a cryogenic probe. The structure was practically determined from the distance, torsion angle, and hydrogen bond restraints listed in Table 1. An overlay of the final 30 structures, which exhibited the best target functions of 100 calculated structures, showed that the coordinates of the backbone atoms were well defined (Fig. 2A). The average root mean square deviations for these structures were 0.63 Å for the backbone (0.25 Å when limited to the $\beta$ strand regions) and 1.19 Å for the side chains. A ribbon representation of the structures of the CNR/Pcdh$\alpha_4$, C-cadherin (Protein Data Bank code 1L3W), N-cadherin (code 1NCI), and E-cadherin (code 1SUH) EC1 domains (displayed with PyMOL). The structural similarity between CNR/Pcdh$\alpha_4$ and each of the other cadherin proteins was estimated by Dali Server Version 2.0 (56). The pairwise root mean square deviations of the C$^\alpha$ atoms and Z-scores obtained were 3.0 Å and 5.8 for C-cadherin, 3.3 Å and 6.0 for N-cadherin, and 3.0 Å and 6.0 for E-cadherin, respectively. The same coloring is used as described for Fig. 2B.
Å for all heavy atoms (0.81 Å for the β strand regions). The detailed statistics for the best 30 structures are also listed in Table 1.

The EC1 domain of CNR/Pcdha4 has the Greek key topology of a β sandwich-like structure containing two β sheets that are packed face to face (Fig. 2). One sheet is composed of four β strands (βA, βC, βF, and βG), and the other is composed of three β strands (βB, βD, and βE). All strands are arranged anti-parallel, except for the parallel pairing between βA and βG. Three loops that are spatially close to the N terminus (loops BC, DE, and FG) and three that are close to the C terminus (loops AB, CD, and EF) connect the corresponding pairs of β strands. The β sandwich scaffold is stabilized by an extensive hydrogen bond network between neighboring β strands and by a hydrophobic core formed by inwardly facing residues from the β sheets (Fig. 2A). The coordinates of the residues forming the hydrophobic core are well defined in the final structures. Of the three aromatic residues (Tyr7, Phe38, and Phe91) contained in the hydrophobic core, Tyr7 is conserved among protocadherin family members.

Despite low sequence similarities between the EC1 domains of CNR/Pcdha4 and the classical cadherins (30% at the maximum), the overall topology of the CNR/Pcdha4 EC1 domain is similar to that of classical cadherin domains in that both contain seven β strands, and the corresponding β strands have similar lengths (Fig. 3). Potentially crucial variations between the structures of the EC1 domains of CNR/Pcdha4 and the classical cadherins are found mainly in the loop regions. Significant differences in the lengths of the loop regions between β strands were demonstrated by aligning the sequences on the basis of the positions of their corresponding β strands (Fig. 3A). First, loop BC of the CNR/Pcdha4 EC1 domain is longer and contains more hydrophobic residues than the domains of classical cadherins, whereas loop FG is shorter. In the homophilic interaction between classical cadherins, these two loops form and surround the hydrophobic pocket that accommodates the N-terminal Try2 residue of the other EC1 domain. Second, loop CD of the CNR/Pcdha4 EC1 domain is much shorter than loop CD of classical cadherins, which has a quasi-β helix conformation. Instead, loop CD of the CNR/Pcdha4 EC1 domain contains the RGD motif, which is known to be a consensus sequence in integrin interactions. Third, a Cys-X5-Cys sequence, which is not found in classical cadherins, is inserted into loop EF of the CNR/Pcdha EC1 domain. Interestingly, we found that the Cys-X5-Cys sequence forms a disulfide bond between Cys70 and Cys76 in loop EF (Fig. 2, B and C). This bond was strongly predicted by the characteristic chemical shift values of the α- and β-carbons of the associated cysteine residues (36) and was confirmed by mass spectrometric analysis of the CNR/Pcdha4 EC1 protein (data not shown).

Calcium Binding Property of the CNR/Pcdha EC1 Domain—Like all classical cadherins, the conserved calcium-binding sites of the CNR/Pcdha EC1 domain (Glu11, Asp65-Arg-Glu67, and Asp99-X-Asn-Asp-Asn109) are clustered near the C terminus, close to the linker region between the EC1 and EC2 domains (Figs. 2B and 4D). A comparison of the 1H, 15N HSQC spectra of the CNR/Pcdha EC1 domain in the presence and absence of calcium ions in solution (1.5 mM) exhibited calcium-induced chemical shift changes in a subset of cross-peaks, including those of Glu11, Arg66, and Asp99. These sites match well with the conserved calcium-binding regions mentioned above. Calcium-induced chemical shift changes were also seen for other residues in the region surrounding these sites (Fig. 4, A and B).

Interestingly, calcium-induced resonance shifts were also observed for the cross-peaks from the Cys70, Gly71, and Ser73 amides forming the disulfide-bonded loop Cys70-Gly-Arg-Ser-Ala-Glu-Cys76 (Fig. 4B). Moreover, the disulfide-bonded loop is spatially located very close to the clusters of calcium-binding residues near the C terminus (Fig. 4D). As the Cys-X5-Cys sequence is conserved among the cluster-type protocadherin families (Pcdha, Pcdhb, and Pcdhγ) (Fig. 4C) and among other non-cluster-type protocadherins (protocadherin-8, Arcadlin), but no in classical cadherins, Cys-X5-Cys may be an additional novel calcium-binding motif unique to the protocadherin family.

Previous analyses of the crystal structures of the EC1 and EC2 domains of E-cadherin (37, 38) and N-cadherin (39) suggested that these calcium-binding residues are involved in calcium-mediated protein-protein interactions. Likewise, the disulfide-bonded loop of the CNR/Pcdha4 EC1 domain may be affected by calcium binding and eventually participate in interactions with other unknown proteins. Generally, the disulfide-bonded loops of Cys-X5-Cys sequences are involved in the protein-protein interactions of viral glycoproteins. For example, a disulfide-bonded loop in the human immunodeficiency virus-1 envelope glycoprotein gp41 plays a central role in gp41-gp120 association and the Env function (40, 41). This example suggests that the Cys-X5-Cys sequence of the CNR/Pcdha family could function as a novel calcium-dependent adhesion interface.

Homophilic Adhesion Interfaces in the Cadherin Superfamily—In the following analysis, we sought to highlight the structural differences between the adhesion interface of classical cadherins and the corresponding region of protocadherins to better understand the adhesion repertoire of the cadherin superfamily. The CNR/Pcdha4 EC1 domain does not have Trp2 or a hydrophobic pocket, which is essential for the adhesiveness of the classical cadherins (Fig. 5A). Among the classical, type II, and desmosomal cadherins, the amino acid residues constitut-
FIGURE 5. Comparison of homophilic binding of the EC1 domains of CNR/Pcdhα and N-cadherin. A, comparison of the adhesive binding sites in the EC1 domain structures in the cadherin superfamily. Shown are backbone structure models viewed from the N terminus and space-filling models viewed from the same direction as in Fig. 3. Residues that form the hydrophobic pocket that is important for adhesiveness in classical cadherins (N-cadherin) and the corresponding cluster of hydrophobic residues in CNR/Pcdhα are highlighted with their side chains shown in red. B, schematic diagram of the chimeric Fc protein variants. The variants include Ncad-Fc (blue diamond), CNREC1/Ncad-Fc (green triangle), CNR/Pcdhα-Fc (red circle), and Fc (black X). sig, signal peptide; PRO, prodomain; TM, transmembrane region; CP, cytoplasmic region. C, Coomassie Blue-stained gel of the purified Fc proteins used in the binding assays (left panel) and beads coated with chimeric Fc protein variants containing similar levels of proteins as assessed by immunoblotting using anti-Fc antibody following the aggregation assay (right panel). Note that the CNR/Pcdhα-Fc fusion protein exhibited almost the same molecular mass as Ncad-Fc and NcadEC1/CNR-Fc, whereas it should be larger by ~100 amino acids. As the CNR/Pcdhα-Fc fusion protein also showed almost the same molecular mass as another classical cadherin (E-cadherin-Fc) in our previous study (22), it appears that classical cadherins undergo more modification (such as glycosylation) than does CNR/Pcdhα. D, time course of bead aggregation monitored by FACS analysis. The percentage of clustered particles is plotted as a function of time. The experiment was performed with at least two different batches of proteins, and the mean S.D. values are shown. E, images of the beads taken at the 120-min time point in D.
ing the hydrophobic pocket (Ile$^{24}$, Tyr$^{36}$, Ala$^{78}$, Ala$^{80}$, and Ile$^{92}$ in the case of N-cadherin) are almost completely conserved (Figs. 3A and 5A). Although the CNR/Pcdhα EC1 domain does have a hydrophobic cluster in the corresponding region, the pocket is not as deep as that in classical cadherins, in which the pocket is large enough to accommodate the side chain of Trp$^2$. The CNR/Pcdhα EC1 domain hydrophobic cluster consists of Ile$^{22}$, Leu$^{26}$, Leu$^{33}$, Phe$^{38}$, Val$^{82}$, Val$^{84}$, and Phe$^{91}$ (Fig. 5A). Among these, Ile$^{22}$, Phe$^{38}$, Val$^{82}$, Val$^{84}$, and Phe$^{91}$ correspond to the amino acids forming the hydrophobic pocket in N-cadherin, respectively. The short side chains of Ala$^{78}$ and Ala$^{80}$ contribute to the formation of the deep hydrophobic space in N-cadherin, whereas the corresponding amino acids in CNR/Pcdhα (Val$^{82}$ and Val$^{86}$) have bulkier side chains that make the cavity smaller. Moreover, some other bulky amino acid residues located in loop BC (Leu$^{26}$, Leu$^{33}$, and Leu$^{23}$) participate in the hydrophobic cluster, filling the hydrophobic cavity and diminishing the open volume of the pocket. Although the calcium-binding sites are conserved in protocadherins as in classical cadherins, the lack of a hydrophobic pocket suggests that the homophilic adhesion interface that is important for classical cadherins does not exist in protocadherins.

These structural differences between the EC1 domains of CNR/Pcdhα and N-cadherin led us to speculate that the previously reported difference in homophilic adhesiveness might be attributable to the difference in EC1 domains alone. We previously detected no homophilic binding activity for the CNR/Pcdhα family proteins in adhesion experiments using protein-coated beads, the design and interpretation of which largely followed those made for classical cadherins (22). To compare the adhesion activity of the EC1 domains more specifically, we replaced the EC1 domain of N-cadherin (22). To compare the adhesion activity of the EC1 domains which largely followed those made for classical cadherins using protein-coated beads, the design and interpretation of aggregation, whereas that of CNR/Pcdhα has six.) However, we observed no significant difference between the aggregation activities of the beads coated with CNR/Pcdhα-Fc and those coated with NcadEC1/CNR-Fc (supplemental Fig. 1). We were not surprised by this result because recent studies demonstrated that the EC1 domain of classical cadherin alone is not sufficient to mediate adhesion activity. Different studies showed that the EC1 and EC2 domains (in a cell aggregation assay) (43) or EC1–EC3 domains (in a bead aggregation assay) (44) are the minimal elements essential for homophilic adhesion activity, even though the EC1 domains are still mainly required because they provide the essential region for the interaction.

These functional data suggest that the difference between the homophilic adhesion of CNR/Pcdhα and that of the classical cadherins is due to the structural differences in the hydrophobic pocket in their EC1 domains. However, we cannot exclude the possibility that structural differences in regions other than the hydrophobic pocket are also partly responsible for lack of CNR/Pcdhα homophilic interaction.

Our structural study of the CNR/Pcdhα EC1 domain and the functional EC1 domain-exchanging experiment provide molecular evidence for differences in the homophilic binding activities of the protocadherins and classical cadherins. The protocadherins may use a different interface, including characteristic loop regions, to exert their functions.

**Heterophilic Adhesion Interface Unique to Protocadherins**—Our solution structure of the CNR/Pcdhα EC1 domain shows that the RGD motif (positions 45–47) is exposed to solvent as a loop. RGD motifs are often found in extracellular proteins as a consensus sequence for interactions with integrins (Fig. 6A). The tenth human fibronectin type III domain of fibronectin also contains the RGD sequence in loop FG, which is known to be its functional binding site for integrin (45). As with the CNR/Pcdhα EC1 domain, the RGD site in the tenth fibronectin type III domain is exposed to solvent as a loop (Fig. 6A). In addition, we determined that the conformation of the RGD site of the CNR/Pcdhα EC1 domain is too flexible to take on a particular secondary structure. This was based on the results of amide$^{15}$N spin relaxation analysis by 500 and 600 MHz NMR assuming anisotropic rotational diffusion (27), which showed the RGD site to exhibit relatively low squared generalized order parameters ($S^2 = 0.63$ for Arg$^{45}$), relatively low $^1$H–$^{15}$N steady-state NOEs (0.62 for Arg$^{45}$ and 0.58 for Asp$^{47}$), and relatively high chemical exchange rates ($R_{ex} = 1.63$ s$^{-1}$ for Arg$^{45}$ and 3.29 s$^{-1}$ for Asp$^{47}$). Generally, low $S^2$ and NOE values indicate the presence of pico- to nanosecond time-scale motions, and high $R_{ex}$ values indicate the presence of micro- to millisecond time-scale motions. Furthermore, the RGD site exhibited amide peaks with very weak intensities in the $^1$H–$^{15}$N HSQC spectra and almost no long-range NOE cross-peaks involving non-labile protons in the two-dimensional NOESY spectra. The solution structure of a mouse fibronectin cell-attachment section consisting of the linked ninth and tenth domain-exchanging experiment (Fig. 5D). As reported previously (42), the ectodomain of full-length N-cadherin induced robust aggregation, whereas that of CNR/Pcdhα showed no aggregation. Notably, the domain-exchanged CNREC1/Ncad-Fc-coated beads exhibited no aggregation. Visualization of the beads by fluorescence microscopy at the 120-min time point revealed aggregates of N-cadherin-coated beads, but not of the other beads, which were not coated with proteins possessing the EC1 domain of N-cadherin (Fig. 5E) (42). As a control experiment, we also exchanged the EC1 domain of CNR/Pcdhα with its counterpart in N-cadherin in the context of CNR/Pcdhα (supplemental Fig. 1). In this experiment, we also removed the EC6 domain of CNR/Pcdhα and constructed NcadEC1/CNR-Fc so that the chimera would have the same number of EC domains as N-cadherin. (N-cadherin has five EC domains, whereas native CNR/Pcdhα has six.) However, we observed no significant difference between the aggregation activities of the beads coated with CNR/Pcdhα-Fc and those coated with NcadEC1/CNR-Fc (supplemental Fig. 1). We were not surprised by this result because recent studies demonstrated that the EC1 domain of classical cadherin alone is not sufficient to mediate adhesion activity. Different studies showed that the EC1 and EC2 domains (in a cell aggregation assay) (43) or EC1–EC3 domains (in a bead aggregation assay) (44) are the minimal elements essential for homophilic adhesion activity, even though the EC1 domains are still mainly required because they provide the essential region for the interaction.

These functional data suggest that the difference between the homophilic adhesion of CNR/Pcdhα and that of the classical cadherins is due to the structural differences in the hydrophobic pocket in their EC1 domains. However, we cannot exclude the possibility that structural differences in regions other than the hydrophobic pocket are also partly responsible for lack of CNR/Pcdhα homophilic interaction.
type III modules, mFnFn3 (9, 10), also shows a disordered RGD site (46). Therefore, the interaction of RGD sites with integrins may be accompanied by an induced fit of the RGD sites. These characteristics of the CNR/Pcdhα4 EC1 domain RGD site meet the two structural criteria required for the biologically active conformation of an RGD site. These criteria are 1) high surface accessibility of the RGD sequence and 2) placement of the RGD site on a loop or a β turn (59).

We recently showed that the RGD motif of the CNR/Pcdhα4 EC1 domain is functionally involved in cell adhesion via β1 integrin (22). Here, we sought to extend this previous functional insight to include the context of the newly determined protein structure. The adhesion capacities of integrins generally require the transition of their structure from a highly bent conformation, which has only low affinities for biological ligands, to an extended conformation through a switchblade-like opening process. The factors known to induce this activation of integrins include divalent cations such as Mn$^{2+}$ in the extracellular medium and several specific “stimulatory” monoclonal antibodies that provoke a conformational change upon binding to a particular integrin (47). To determine whether the conformational activation of integrin is required for it to bind the CNR/Pcdhα EC1 domain, we modified our previous cell adhesion assay from one using serum in the cell suspension buffer, which may cause the nonspecific activation of integrins (22), to one using serum-free conditions.
FIGURE 7. Evolutionary features of the CNR/Pcdhα EC1 domain structure. A, quality index scores (QI) of multiple alignments of the CNR/Pcdhα family members from various species. Note that, in the zebrafish EC1 domains, the sequences are greatly diversified. B, the degree of EC1 domain sequence conservation for each species mapped using the Consurf server (58). Paralogously conserved residues are mapped on the structures of CNR/Pcdhα. Highly conserved residues are in red, and diversified residues are in blue.
using Mn²⁺ or the β1 integrin stimulatory monoclonal antibody TS2/16 (48) to induce directly the activation of β1 integrins. In the revised assay system, HEK293T cells exhibited no adhesion to wells coated with CNR/Pcdha4 EC1-Fc protein in the absence of manganese (1 mM MgCl₂ and CaCl₂) (data not shown). However, the addition of manganese (1 mM MnCl₂) caused a dramatic rise in the adhesion activity of the HEK293T cells (Fig. 6, B, panels a and C). The adhesion efficacy of CNR/Pcdha4 EC1-Fc was similar to that of CNR/Pcdha-Fc (supplemental Fig. 2), as reported using the previous assay system. Moreover, the specific activation of β1 integrin by antibody TS2/16 also increased the cell adhesion to CNR/Pcdha4 EC1-Fc, even in the absence of manganese (1 mM MgCl₂ and 1 mM CaCl₂). Control IgG induced no adhesion (Fig. 6, B, panels b–d; and C). The activation with antibody TS2/16 was specific because the control Fc protein did not exhibit cell adhesion (Fig. 6, B, panel e; and C). These data indicate that the activation of β1 integrin is at least partly required for the binding of HEK293T cells to CNR/Pcdha4 EC1-Fc. Although our present and previous data do not exclude the possibility of other sites outside the RGD being involved in the interaction, these structural and functional data suggest that the RGD loop of the CNR/Pcdha4 EC1 domain acts as at least one of the binding sites for integrins in the active conformation.

Evolutionary Features of the Structure of the CNR/Pcdha Family EC1 Domain—One feature of the CNR/Pcdha family is its molecular diversity among species (18, 49–53). Although the EC1 domain of CNR/Pcdha is conserved in terms of its sequence among family members in mouse, rat, chick, and human, the EC1 domain of the zebrafish CNR/Pcdha differs significantly among family members (Fig. 7A). To assess the significance of this diversity, we mapped the information from multiply aligned sequences for each species onto the CNR/Pcdha4 EC1 domain structure (Fig. 7B). In every species, most of the sequentially diverse regions are restricted to the surface area of the structure. In particular, the amino acid sequences of the disulfide-bonded loops between Cys⁷⁰ and Cys⁷⁶ tend to be diverse. In the case of zebrafish, there is extensive diversity over the entire molecular surface. The conservation of the EC1 domain in mammals, in contrast to the diversification of this domain in zebrafish, may reflect different roles played by the CNR/Pcdha family in the markedly different brain structures of different vertebrate classes.

Conclusion—We have determined, for the first time, the structure of a protein in the protocadherin family, which accounts for the largest subgroup within the cadherin superfamily. Several characteristic features were revealed in the structure of the EC1 domain for this family. These include (i) the lack of an interface for the homophilic adhesiveness that is typically found in classical cadherins, (ii) the loop region structures as molecular interfaces distinct from those of classical cadherins, and (iii) diversity in the surface structures relating to evolutionary differences.

For years, our understanding of the adhesion repertoire of the cadherin superfamily was derived largely from studies of the classical cadherin subfamily. Our analysis of the much larger protocadherin subfamily extends this understanding, indicating that the adhesion repertoire of this superfamily is greater than previously expected. Moreover, these results provide a framework for increasing our understanding of the functions of the cadherin superfamily. Because the members of the protocadherin family are expressed mainly in the nervous system, their diverse structures and functional repertoire may be specifically required for the construction of the highly organized brain. Our study also paves the way for studies designed to reveal more about the molecular basis of the extraordinary diversity of brains.

Acknowledgments—We thank Woj M. Wojtowicz and Dr. Lawrence Zipursky (UCLA, Los Angeles, CA) for the pLb/Fc vector, Woj M. Wojtowicz for useful suggestions for setting up the bead aggregation assay, Dr. Rene Marc Mege (INSERM, Paris) for the pCEP4-CECH vector, Tomoyo Takai and Tomonori Utsuyama for support during NMR analyses, Seisuke Kusano (University of Hyogo) for supporting the initial phase of this work, Dr. Kiyotoshi Sekiguchi and Toshihiko Uemura for antibody TS2/16 and useful discussions, Dr. Rie Nagao-Kiyama for supporting functional analysis, Dr. Shun Hamada (Fukuoka Women’s University) for valuable suggestions and critical reading of the manuscript, and the members of the Yagi and Ikegami laboratories for assistance and numerous discussions.

REFERENCES

1. Takeichi, M. (1988) Development (Camb.) 102, 639–655
2. Tepass, U., Truong, K., Godt, D., Ikura, M., and Pfeifer, M. (2000) Nat. Rev. Mol. Cell Biol. 1, 91–100
3. Yagi, T., and Takeichi, M. (2000) Genes Dev. 14, 1169–1180
4. Nollet, F., Kools, P., and van Roy, F. (2000) J. Mol. Biol. 299, 551–572
5. Overduin, M., Harvey, T. S., Bagby, S., Tong, K. I., Yue, P., Takeichi, M., and Ikura, M. (1995) Science 267, 386–389
6. Shapiro, L., Kwong, P. D., Fannon, A. M., Colman, D. R., and Hendrickson, W. A. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 6793–6797
7. Shapiro, L., Fannon, A. M., Kwong, P. D., Thompson, A., Lehmann, M. S., Grubel, G., Legrand, J. F., Als-Nielsen, J., Colman, D. R., and Hendrickson, W. A. (1995) Nature 374, 327–337
8. Shapiro, L., and Colman, D. R. (1998) Curr. Opin. Neurobiol. 8, 593–599
9. Shapiro, L., and Colman, D. R. (1999) Neuron 23, 427–430
10. Boggan, T. J., Murray, J., Chappuis-Flament, S., Wong, E., Gumbiner, B. M., and Shapiro, L. (2002) Science 296, 1308–1313
11. Patel, S. D., Chen, C. P., Bahna, F., Honig, B., and Shapiro, L. (2003) Curr. Opin. Struct. Biol. 13, 690–698
12. Gooding, J. M., Yap, K. L., and Ikura, M. (2004) BioEssays 26, 497–511
13. Koch, A. W., Manzur, K. L., and Shan, W. (2004) Cell. Mol. Life Sci. 61, 1884–1895
14. Frank, M., and Kemler, R. (2002) Curr. Opin. Cell Biol. 14, 557–562
15. Suzuki, S. T. (2000) Exp. Cell Res. 261, 13–18
16. Wu, Q., and Maniatis, T. (1999) Cell 97, 779–790
17. Sugino, H., Hamada, S., Yasuda, R., Fuji, A., Matsuda, Y., Fujita, M., and Yagi, T. (2000) Genomics 63, 75–87
18. Wu, Q., Zhang, T., Cheng, J. F., Kim, Y., Grimwood, J., Schmutz, J., Dickson, M., Noonan, J. P., Zhang, M. Q., Myers, R. M., and Maniatis, T. (2001) Genome Res. 11, 389–404
19. Kohmura, N., Senzaki, K., Hamada, S., Kai, N., Yasuda, R., Watanabe, M., Ishii, H., Yasuda, M., Mishina, M., and Yagi, T. (1998) Neuron 20, 1137–1151
20. Tasic, B., Nabholz, C. E., Baldwin, K. K., Kim, Y., Rueckert, E. H., Ribich, S. A., Cramer, P., Wu, Q., Axel, R., and Maniatis, T. (2002) Mol. Cell 10, 21–33
21. Esumi, S., Kazaku, N., Taguchi, Y., Hirayama, T., Sasaki, A., Hiraibayashi, T., Koide, T., Kitsuikawa, T., Hamada, S., and Yagi, T. (2005) Nat. Genet.
37. Mutoh, T., Hamada, S., Senzaki, K., Murata, Y., and Yagi, T. (2004) *Exp. Cell Res.* **294**, 494–508.

22. Tsumoto, K., Shinoki, K., Kondo, H., Uchikawa, M., Juji, T., and Kumagai, I. (1998) *J. Immunol. Methods* **219**, 119–129.

23. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., III, and Skelton, N. J. (1996) *Protein NMR Spectroscopy*, Academic Press, San Diego, CA.

24. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR* **6**, 277–293.

25. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) *Biochemistry* **33**, 5984–6003.

26. Mandel, A. M., Akke, M., and Palmer, A. G., III (1995) *J. Mol. Biol.* **246**, 144–163.

27. Cornilescu, G., Delaglio, F., and Bax, A. (1999) *J. Biomol. NMR* **13**, 289–302.

28. Wuchter, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, Inc., New York.

29. Holm, L., and Sander, C. (1993) *J. Mol. Biol.* **233**, 123–138.

30. Page, R. D. (1996) *Comput. Appl. Biosci.* **12**, 357–358.

31. Aiyar, A. (2000) *Methods Mol. Biol.* **132**, 221–241.

32. Sayle, R. A., and Milner-White, E. J. (1995) *Trends Biochem. Sci.* **20**, 374.

33. Armon, A., Graur, D., and Ben-Tal, N. (2001) *J. Mol. Biol.* **307**, 447–463.

34. Torshin, I. Y. (2002) *Med. Sci. Monit.* **8**, 301–312.