Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 6793–6797, July 1995
Biochemistry

Considerations on the folding topology and evolutionary origin of cadherin domains

LAWRENCE SHAPIRO*, PETER D. KWONG*, ALLISON M. FANNON†, DAVID R. COLMAN†, AND WAYNE A. HENDRICKSON*‡

*Department of Biochemistry and Molecular Biophysics and ‡Howard Hughes Medical Institute, Columbia University, New York, NY 10032; and †Brookdale Center for Molecular Biology, Mount Sinai School of Medicine, New York, NY 10029

Contributed by Wayne A. Hendrickson, March 7, 1995

ABSTRACT **Cell–cell adhesion in zonula adherens and desmosomal junctions is mediated by cadherins, and recent crystal structures of the first domain from murine N-cadherin provide a plausible molecular basis for this adhesive action. A structure-based sequence analysis of this adhesive domain indicates that its fold is common to all extracellular cadherin domains. The cadherin folding topology is also shown to be similar to immunoglobulin-like domains and to other Greek-key β-sandwich structures, as diverse as domains from plant cytochromes, bacterial cellulases, and eukaryotic transcription factors. Sequence similarities between cadherins and these other molecules are very low, however, and intron patterns are also different. On balance, independent origins for a favorable folding topology seem more likely than evolutionary divergence from an ancestor common to cadherins and immunoglobulins.**

The cadherins are a large class of calcium-dependent cell–cell adhesion molecules (1, 2). They are single-pass transmembrane proteins characterized by distinctive extracellular sequence repeats of ≈110 aa (3). Cadherins are homophilic; each cadherin subtype specifically interacts with like molecules presented on neighboring cells. The determinants of adhesive specificity are thought to reside in the N-terminal repeat (4). The prototypical cadherins are those found in the zonula adherens cell–cell junctions of vertebrates. These "classic" cadherins have five cadherin repeat sequences in their extracellular segments and a cytoplasmic domain that is specialized to mediate connections with cytoskeletal actin filaments. Regulated expression of the classic cadherins is important in many morphogenic processes. Another subfamily of cadherin molecules derives from desmosomal junctions. These proteins are similar to the classic cadherins, but their cytoplasmic domains are specialized to interact with a plaque ultimately connected to intermediate filaments (5). Cadherin-like proteins have also been identified in *Drosophila* (2, 6), and sequences resembling the cadherin repeat motif are also found in other proteins (7).

Our recent crystal structures of the adhesive domain of mouse N-cadherin (N-cadherin domain 1, NCD1) reveal a zipper-like superstructure of protomers that appears to be relevant for cell–cell adhesion by cadherins (4). The protomers of this zipper make up compact β-sandwich domains. Two repeated dimer interfaces form the zipper. One generates dimers in a head-to-head orientation and corresponds to the adhesive interface formed between individual cadherin molecules from opposing cell surfaces. The other joins protomers in a parallel orientation, as in dimers emanating from the same cell surface. Analogs of Ca^{2+} (Yb^{3+} and UO_2^{2+}) are ligated between two loops near the C terminus, away from the adhesive interface, at a proposed site of interdomain Ca^{2+} coordination (4). After we had completed our NCD1 structure analysis, an NMR study of the N-terminal domain of E-cadherin was published (8); the reported structure is similar to NCD1 in its overall fold, but it is monomeric.

Here we present a detailed structural characterization of NCD1, discuss the probable generality of the cadherin fold, compare the NCD1 structure with some other β-sandwich proteins, and analyze the evolutionary implications of these findings.

METHODS

Structural Analysis. Atomic models of NCD1 were taken from refinements of the structure in three lattices as described (4). There are five crystallographically distinct copies: one in the type A ($P6_322$) lattice (R = 0.217 at 1.9-Å resolution, T = −160°C), two in the type B ($P321$) lattice (R = 0.224 at 2.1-Å resolution, T = 20°C), and two in the type C ($P2_12_12$) lattice (R = 0.195 at 1.9-Å resolution, T = −160°C). Except where otherwise noted, type C molecule 1 has been used. Secondary structural elements were assigned by the method of Kabsch and Sander (9). Main-chain hydrogen bonds were assigned for Co···N distances <3.5 Å and CO···N angles >120°. Fractional solvent accessibilities were defined as compared with the corresponding Gly-Xaa-Gly peptide (10).

Molecular Superpositions. Cadherin domain NCD1 was compared with other Greek-key β-sandwich proteins: the light-chain variable region (V_L) and heavy-chain constant region 1 of Fab New (11), D1 of CD4 (12), telokin (13), limulus hemocyanin domain 3 (14), neocarzinostatin (15), and the large subunit of cytochrome *f* (16), from the Protein Data Bank (7fab, 1cdh, 1tlk, 1lla, 1noa, and 1ctm, respectively); D1 from CD2 (17, 18) and D1 from vascular cell adhesion molecule 1 (VCAM-1) (19), from Yvonne Jones; and the p50 subunit of NF-κB (20) from Greg Van Duyne. Molecules were aligned visually in INSIGHT, extended by the procedure of Rossmann and Argos (21), and finally optimized with TOSS (22) to include segments with at least three contiguous Cα positions all within 2.5 Å.

Structure-Based Sequence Alignments. Obviously homologous sequences (e.g., cadherin domains) were aligned to the NCD1 domain with gaps restricted to regions between elements of secondary structure. In cases of very low similarity, sequences were aligned in accord with structural superpositions, again restricting gaps to interstrand segments. Gap positions were adjusted to maximize the pairwise sequence identities in each comparison.

RESULTS AND DISCUSSION

Structural Characteristics of the Adhesive Domain of N-Cadherin. As described (4), NCD1 has seven β-strands arranged in two β-sheets such that the N and C termini emanate from opposite ends of the sandwich. One edge of the sandwich features a quasi-β-helix and the other is engaged in a strand-

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

Abbreviations: V_L, light-chain variable region; VCAM-1, vascular cell adhesion molecule 1; NCD1, N-cadherin domain 1.

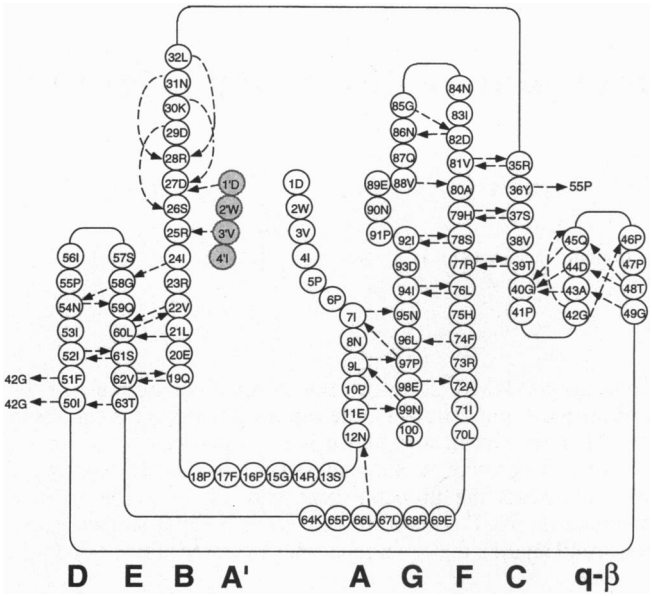

FIG. 1. Main-chain hydrogen bonding pattern for NCD1. The arrows represent hydrogen bonds and point from the N—H donor to the C=O acceptor.

mediated dimer interface. The hydrogen-bonding pattern that defines this folding topology is shown in Fig. 1. The β-strands have been labeled strands A–G with spans of 1–10, 19–26, 35–39, 51–54, 59–63, 74–82, and 92–100 aa, respectively, as defined algorithmically (9) in the context of the strand dimer interface. The only other regular feature is a 3_{10}-helix spanning aa 26–31. The quasi-β-helix is so named because of a β-like hydrogen-bonding pattern and an overall helical appearance. It is, however, irregular with special conformations only accessible to Gly residues at positions 40, 42, and 49 and a cis Pro residue at position 47. The quasi-β-helix is involved in the adhesion interface. Gly-58 and Gly-85 specify unique conformations in the DE and FG loops, which are also important in adhesion. Two additional cis Pro residues are found at positions 16 and 18 in the AB loop. Otherwise all peptides are trans and all residues have fully allowed conformations.

The structure of NCD1 has been determined in five distinct environments in three lattices, and there is a small but significant variability in these structures (Fig. 2*A*). The rms deviation from the mean in a superposition of backbone (N, Cα, Cβ, C, and O) atoms is 0.39 Å overall (aa 1–100) and this reduces to 0.24 Å for β-strands as defined above. Deviations exceed 1.9 Å between type A and type C structures for aa 67–69, which differ due to binding of Yb^{3+} and UO_2^{2+}, respectively. Deviations from the mean also exceed 1 Å for aa 1–5. This, however, is due to flexion of the strand dimer; these residues are similar (0.50 Å, rms deviation) in the hydrogen-bonded context of the strand dimer partner. Another region of variability is in the BC loop, which in succeeding domains contains Ca^{2+} ligands. Moreover, aa 101–108, which include the conserved DXNDN sequence also thought to be involved in Ca^{2+} coordination, are completely disordered in these crystals. Thus, conformational adaptability is expected in the interdomain calcium coordination that we have proposed (4). Elevated deviations also exist in the AB, EF, and FG loops, and these results are generally consistent with atomic mobility patterns as revealed by main-chain B factors (Fig. 3).

Fractional solvent accessibility values are given in Table 1 for an isolated protomer and also in the context of the adhesive zipper interfaces. Fig. 2*B* shows side chains that are buried, both those in the isolated domain and those covered in the dimer interfaces. It is especially notable that Trp-2′ (prime indicating a partner molecule) is an integral part of the hydrophobic core.

Generality of the Cadherin Fold. A structure-based alignment of sequences (Fig. 4) suggests to us that all five extracellular cadherin repeats share a common folding topology. The similarity is greatest between corresponding domains across the family (e.g., 58% identity between D1 of N- and E-cadherin), but it is also significant among domains D1 to D4 within a molecule (21–32% for murine N-cadherin). D5 is less related to the others (10–16% identity), but notable features are preserved. In particular, as discussed (4), residues proposed for interdomain Ca^{2+} coordination are conserved. While no residues are identical among all cadherin domains, a core of large hydrophobic residues in NCD1 (including Trp-2′, Tyr-36, Phe-51, and Phe-74) is preserved in character across the family. Positions of insertions and deletions are conserved, and all are within domains (Fig. 4).

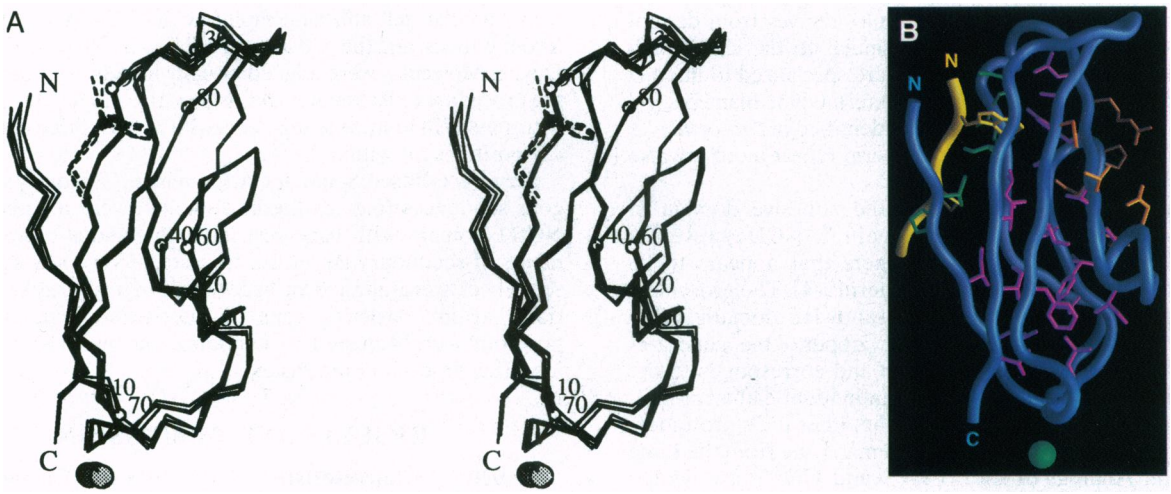

FIG. 2. Structure of the cadherin domain NCD1. (*A*) Cα superposition of the five replicates of NCD1 in the three crystal forms. Cα positions for aa 1′–5′ and the side chains of the Trp-2′ are shown as dotted lines for the strand dimer-related partner molecules. The Yb^{3+} of the type A and B crystals are drawn as dark spheres, and the uranium atom of the uranyl ion of the type C crystals is drawn as a lighter sphere. (*B*) Backbone worm diagram of NCD1. The polypeptide backbone is blue, and buried side chains (fractional solvent accessibility <10%) in the isolated NCD1 protomer are pink. A backbone trace through aa 1–7 of the strand dimer-related molecule together with its side chains that are buried in the strand dimer interface, but not in the isolated protomer, are yellow. Side chains of the main protomer that are buried in the strand dimer interface are green, and side chains buried in the cell-adhesion interface are orange.

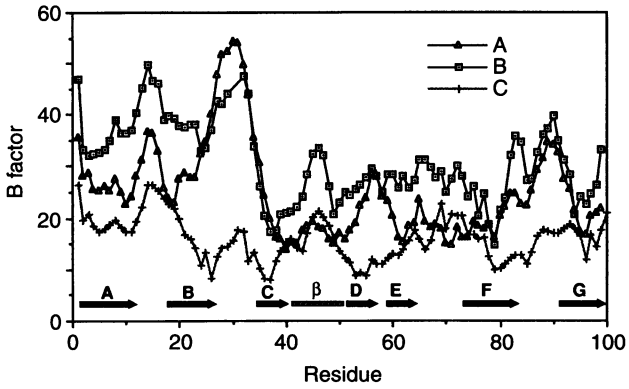

FIG. 3. Atomic mobility in NCD1 crystals. Average main-chain (N, Cα, Cβ, C, and O) B factors are shown from the type A (△), type B copy 1 (□), and type C copy 1 (+) structures.

Conservation among D1 sequences is especially notable in the strand dimer interface. The interaction of Trp-2′ from a partner molecule with the hydrophobic interior of the domain dominates the interface, and the contacted groups are highly conserved. Thus, Tyr-36 and Ala-80 are always the same, and others are highly restricted: Ile/Val-24, Ala/Ser/Cys-78, Ile/Met/Leu-92, and Ile/Val-96. Glu-89, which interacts with the partner N terminus, is also invariant. Phe residues at positions in domains D2–D4 that are analogous with Trp-2 in D1 may also intercalate into strand dimer partners (4). The prospective contacting groups are conserved in character in these domains as well, but alternatives to strand exchange cannot be excluded. D2–D5 are not expected to have quasi-β-helices.

In keeping with this anticipated generality of the cadherin fold, the N-terminal domain structure of murine E-cadherin reported from an NMR analysis (8) is similar in overall fold to the D1 structure described here for N-cadherin. Apart from relatively minor distinctions (e.g., we see no evidence for an α-helix in the EF loop), the major difference relates to positioning of the A strand. The monomeric NMR structure shows the beginning of the A strand hydrogen bonded to its own B strand, with the side chain of Trp-2 exposed. These differences are surprising in light of the sequence similarities between these molecules generally (58% identity in D1 domains) and in the strand dimer interface particularly (only Ile/Met-92 differ for the indole ring contacts in NCD1).

In addition to their defining occurrence in the classic and desmosomal cadherins, other cadherin domains are found in a *Drosophila* cadherin, which has extra repeats (2), in the *fat* tumor suppresser gene of *Drosophila* (6), which contains 34 tandem cadherin repeats, and in the protooncogene c-*ret* (7).

Table 1. Fractional solvent accessibility

Residue	% A	% B	Residue	% A	% B	Residue	% A	% B
1 Asp	49	(26)	34 Leu	6		67 Asp	38	
2 Trp	78	(0)	35 Arg	53	[4]	68 Arg	33	
3 Val	69	(24)	36 Tyr	1	(0)	69 Glu	39	
4 Ile	19	(1)	37 Ser	24	[19]	70 Leu	79	
5 Pro	68	(5)	38 Val	8		71 Ile	33	
6 Pro	45	(34)	39 Thr	26	[6]	72 Ala	29	
7 Ile	11		40 Gly	8	[0]	73 Arg	50	
8 Asn	48		41 Pro	33		74 Phe	1	
9 Leu	0		42 Gly	0		75 His	50	
10 Pro	49		43 Ala	7		76 Leu	1	
11 Glu	11		44 Asp	52	[9]	77 Arg	40	[39]
12 Asn	50		45 Gln	41	[31]	78 Ala	3	(0)
13 Ser	14		46 Pro	68		79 His	16	[4]
14 Arg	90		47 Pro	37		80 Ala	3	(0)
15 Gly	21		48 Thr	53	[47]	81 Val	21	[14]
16 Pro	79		49 Gly	40	[35]	82 Asp	17	[16]
17 Phe	26		50 Ile	1		83 Ile	87	[48]
18 Pro	61		51 Phe	1		84 Asn	76	[30]
19 Gln	27		52 Ile	42	[6]	85 Gly	52	[7]
20 Glu	55		53 Ile	6	[2]	86 Asn	61	[30]
21 Leu	19		54 Asn	37	[16]	87 Gln	45	[36]
22 Val	31	(13)	55 Pro	34	[0]	88 Val	17	
23 Arg	47	(38)	56 Ile	73	[49]	89 Glu	30	(4)
24 Ile	3	(0)	57 Ser	24		90 Asn	63	(55)
25 Arg	33	(9)	58 Gly	0		91 Pro	21	[17]
26 Ser	18	(0)	59 Gln	18		92 Ile	26	(3)
27 Gly	52	(17)	60 Leu	0		93 Asp	38	
28 Arg	41	(24)	61 Ser	8		94 Ile	0	
29 Asp	19		62 Val	0		95 Val	15	
30 Lys	74		63 Thr	26	[23]	96 Ile	1	
31 Asn	74		64 Lys	49		97 Asn	13	
32 Leu	44		65 Pro	41		98 Val	0	
33 Ser	64		66 Leu	7		99 Ile	42	
						100 Asp	74	

Fractional solvent accessibility is shown for each residue as a percentage, calculated as described in the text. Column A gives values for an isolated monomer; column B gives values in the context of the interfaces of the adhesive ribbon, with those altered by strand dimer contacts shown in parentheses and those from the adhesion dimer shown in square brackets.

The similarity in c-*ret* appears to begin at the top of the C-strand and continues through strands D, E, F, and G and then to strands A and B of a connected cadherin domain. Large insertions may have obscured other strand identifications.

Comparison with Other Greek-Key β-Sandwiches. The folding topology of NCD1 is reminiscent of the immunoglob-

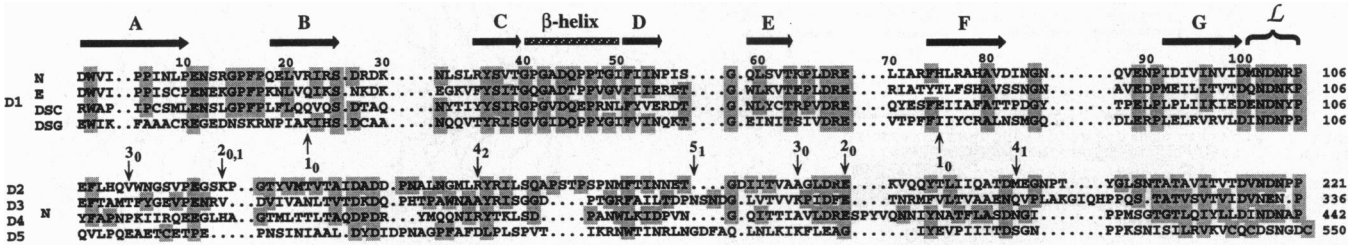

FIG. 4. Sequence alignments of murine N-cadherin (N) with human E-cadherin (E), human desmocollin 2 (DSC), and human desmoglein I (DSG). The upper grouping shows sequences of the D1 domains of these molecules. These have no insertions or deletions; gaps have been inserted to facilitate alignment with the non-D1 domains below. The second group shows the aligned sequences of D2–D5 of murine N-cadherin. Aligned residues that match in corresponding domains of three of the four proteins are shaded for both groups. A gap is placed in the A strand of D1 domains to facilitate alignment of the conserved Phe residues of D2–D4 with Trp-2 of D1 domains. Otherwise, gaps are restricted to between strands. Vertical arrows mark the exon boundaries. Numbers associated with these arrows give the affected domain number, while the subscripts indicate the phase of the exon boundary; the last amino acid coded by an exon in N-cadherin is marked. Assignment of the β-strand positions was by the Kabsch and Sander algorithm (9) for all but strands A and B, which have been extended to include β-type hydrogen bonding across the strand dimer interface. ℒ marks the segment that links successive cadherin domains.

ulin superfamily (Fig. 5*A*). The cadherin fold is more similar to immunoglobulin V domains than to constant domains, taking placement of the A strand and AB loop as defining characteristics (23), but it is constant-like in the absence of C′ and C″ strands (replaced in NCD1 by the quasi-β-helix). The structures of NCD1 and several immunoglobulin-like domains have been superimposed and used to produce sequence alignments (Fig. 5 and Table 2). The structure-based sequence matches of NCD1 with these and other V-like domains (including Rei, CD2, CD4 D3, and CD8) are extremely low at <11% identity. In contrast, the tertiary structures superimpose fairly well although most loop regions are highly divergent. In particular, intersheet strand connections AB and EF at the C-terminal end of NCD1 are much broader than their counterparts in immunoglobulin-like domains.

A number of other protein domains are folded as Greek-key β-sandwiches (24) with an immunoglobulin-like topology. This topological similarity was first noted for superoxide dismutase (25) and was later extended to a domain from arthropod hemocyanins and the neocarzinostatin family of antitumor proteins (26) and also in the bacterial proteins PapD (27) and cellulase (28). Recently, such folds have also been found in the p53 tumor suppressor protein (29), the large subunit of cytochrome *f* (16), and in transcription factor NF-κB (20, 30). We have compared both NCD1 and the V_L domain from IgG New, a representative immunoglobulin domain, with several of these other Greek-key β-sandwich proteins (Table 2). The structural superpositions for some of these, notably for NCD1 with D1 of NF-κB (Fig. 5*C*) and with the large domain of cytochrome *f*, are quite close and extensive. The common core in all these proteins contains strands B and E from one sheet and strands C, F, and G from the other. Variations occur in the positioning of edge strands A and D(C′), which can be in either of the two sheets or switch between them. Structural similarities are greatest within common topological subsets, as for N-cadherin D1 and VCAM D1 (Fig. 5*C*).

Comparison of Adhesive Interfaces. To our knowledge, the structure of NCD1 provides the first view of a specific cell-adhesive interface. However, crystal structures of the immunoglobulin superfamily cell adhesion molecule CD2, which is not known to function homophilically, show crystal contacts that may illustrate the mode of CD2 binding to its natural ligand CD58 (LFA-3) (17, 18). Although the NCD1 and CD2 interfaces share some similarities, the differences loom larger. The CD2 contact is mediated entirely through the GFCC′C″ sheet, with the primary interactions found between the FG and C′C″ loops. In NCD1 the C′ and C″ strands are replaced by the quasi-β-helix, but interactions are with a symmetry mate of itself rather than with the partner FG loop, which instead interacts with the DE. With one molecule of the NCD1

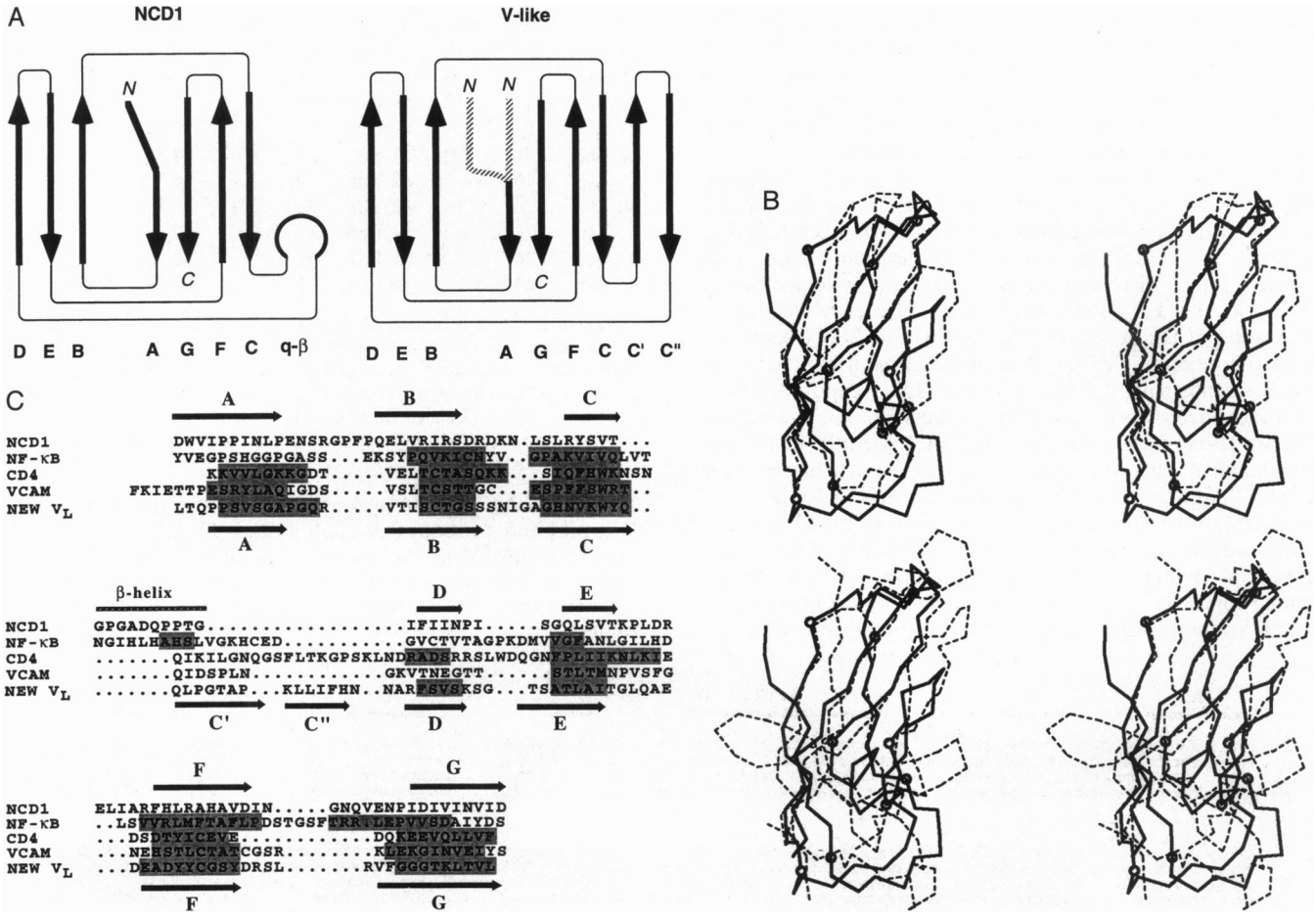

FIG. 5. Comparison of NCD1 with representative β-sandwich domains. (*A*) Topology diagrams for NCD1 and immunoglobulin V-like domains. The first half of the V-like A strand has alternative placements (hatched region), while that for NCD1 veers away into the dimer interface with a partner molecule. The quasi-β-helix region of NCD1 is represented by a circular loop. A short 3_{10}-helix extending from aa 27 to 30 is found in the BC loop, which is highly mobile. (*B*) Stereo diagram of domain 1 from VCAM-1 REI (*Upper*) and NF-κB (*Lower*) in dotted lines superimposed on NCD1 in solid lines. Cα traces are shown, with spheres drawn for every tenth Cα of NCD1. For clarity, aa 142–205 of the EF loop in NF-κB are not shown. (*C*) Structure-based alignment of the amino acid sequences of NCD1, the V_L domain from IgG New, D1 of VCAM-1, D1 of CD4, and NF-κB. Shaded regions have contiguous Cα positions within 2.5 Å from the corresponding NCD1 positions. Secondary structure positions are shown for NCD1 above the sequences and for CD4 below. Alignments in loop regions for the multiple sequences shown here may differ from those in pairwise alignments (Table 2).

Table 2. Structural and sequence relationship between some β-sandwich domains

	Sequence identity, %		Structural superposition	
	NCD1	V_L New	NCD1	V_L New
V_L (IgG New)	10	—	1.76 (51)	—
CD4 D1	6	20	2.12 (55)	0.94 (60)
VCAM D1	8	15	1.35 (49)	1.34 (63)
Telokin	11	13	1.31 (41)	1.11 (68)
CH1 (IgG New)	8	9	1.57 (28)	1.61 (33)
FNIII (tenascin)	9	6	1.49 (38)	1.39 (31)
Neocarzinostatin	13	12	2.01 (30)	1.82 (29)
Cytochrome *f*	10	5	1.53 (42)	1.41 (51)
NF-κB D1	8	9	1.51 (42)	1.39 (38)
Cadherin (NCD1)	—	10	—	1.76 (51)

Sequence identity percentages are from structure-based alignments, as in Fig. 5, with gaps given one place in the denominator. Structural superposition values give rms deviations between Cα positions for segments that match within 2.5 Å with the number of matches in parentheses. CH1, heavy-chain constant region 1; FNIII, type III fibronectin domain.

adhesion dimer superimposed on a molecule of the CD2 lattice contact, the partner NCD1 molecule is rotated by 83° (for human CD2) to 92° (for rat CD2 molecule 2) from the interface-related CD2 molecule.

Evolutionary Origins. The unexpected similarity in folding topology between cadherin and immunoglobulin domains suggests the possibility of a common ancestor. This alone has led some to conclude that cadherins belong to the immunoglobulin superfamily (8, 31), and instances of rather good superimposability (Fig. 5*C* and Table 2) would seem to strengthen this conclusion. On the other hand, the Greek-key β-sandwich that is at the core of this folding pattern recurs in diverse circumstances. Thus, others infer that such proteins have a common intrinsic folding propensity but distinct evolutionary origins (26).

If cadherins and immunoglobulin-like molecules do derive from a common origin, the point of divergence must be remote. Sequence similarities between the two types of domains are vanishingly small (6–11% identity, Table 2). This is despite a rather slow evolutionary rate for cadherins; sequences of mouse and human N-cadherins are 96% identical (32) and the domains from *Drosophila* (2, 6) and vertebrate cadherins are typically 24–35% the same. In contrast, the immune system is a relatively recent adaptation, by vertebrates, and yet V and constant domains from the family are only 9% identical and immunoglobulin-like cell adhesion domains are only 15–20% identical (Table 2). Moreover, whereas sequence similarity and structural deviations are correlated in the globin family (33), the remarkably good superposition of the cadherin domain on some immunoglobulin-like domains has no associated basis in sequence (Table 2).

Intron patterns can also provide clues to evolution. Thus, repeated immunoglobulin or type III fibronectin domains are commonly separated by introns that split the reading frame in phase 1 (34), indicating tandem duplication by class one to one recombination. The intron structure of cadherin genes is distinctly different (ref. 35 and references therein). Here all introns are within rather than between domains and are primarily in phase 0 (Fig. 4). The conserved intron pattern of cadherins implies that domain duplication preceded the division into tissue- and junction-specific types, perhaps by a mechanism different from exon shuffling.

The alternative to a remote divergence is an independent initiation of the Greek-key topology that underlies the folds of cadherin and immunoglobulin domains. This topology is distributed among such functionally distinct and phylogenetically disparate molecules as vertebrate immune system proteins, plant cytochromes, bacterial cellulases and molecular chaperones, and eukaryotic transcription factors and cell adhesion molecules. Many of these may have evolved in parallel—neither diverging from a common ancestor nor converging onto a common function. It may simply be an energetically favorable and kinetically readily accessible folding that is in common. Generally accepted examples range from α-helices themselves to commonplace folding topologies such as four-helix bundle proteins of the myohemerythrin topology and α/β-barrel proteins of the triose-phosphate isomerase topology (36). The distinctive hydrophobic core signatures of cadherin (4), fibronectin type III (37), and immunoglobulin modules are compatible with independent origins of this kind.

In conclusion, while a common evolutionary origin cannot be excluded, independent derivations for this frequently occurring topology seems more likely.

We thank Yvonne Jones, Greg van Duyne, and Gouri Ghosh for generously providing coordinate sets. This work was supported in part by National Institutes of Health grants (D.R.C.), a National Eye Institute training grant (L.S.), and Human Frontiers Science and National Multiple Sclerosis Society fellowships (A.M.F.).

1. Takeichi, M. (1990) *Annu. Rev. Biochem.* **59,** 237–252.
2. Oda, H., Uemura, T., Harada, Y., Iwai, Y. & Takeichi, M. (1994) *Dev. Biol.* **165,** 716–726.
3. Hatta, K., Nose, A., Nagafuchi, A. & Takeichi, M. (1988) *J. Cell Biol.* **106,** 873–881.
4. Shapiro, L., Fannon, A. M., Kwong, P. D., Thompson, A., Lehmann, M. S., Grübel, G., Legrand, J.-F., Als-Nielsen, J., Colman, D. R. & Hendrickson, W. A. (1995) *Nature (London)* **374,** 327–337.
5. Magee, A. I. & Buxton, R. S. (1991) *Curr. Opin. Cell Biol.* **3,** 854–861.
6. Mahoney, P. A., Weber, U., Onofrechuk, P., Biessmann, H., Bryant, P. J. & Goodman, C. S. (1991) *Cell* **67,** 853–868.
7. Schneider, R. (1992) *Trends Biochem. Sci.* **17,** 468–469.
8. Overduin, M., Harvey, T. S., Bagby, S., Tong, K. I., Yau, P., Takeichi, M. & Ikura, M. (1995) *Science* **267,** 386–389.
9. Kabsch, W. & Sander, C. (1979) *Biopolymers* **22,** 2577–2637.
10. Shrake, A. & Rupley, J. A. (1973) *J. Mol. Biol.* **79,** 351–371.
11. Saul, F. A., Amzel, L. M. & Poljak, R. J. (1978) *J. Biol. Chem.* **253,** 585.
12. Ryu, S.-E., Truneh, A., Sweet, R. W. & Hendrickson, W. A. (1994) *Structure* **2,** 59–74.
13. Holden, H. M., Ito, M., Hartshornr, D. J. & Rayment, I. (1992) *J. Mol. Biol.* **277,** 840–851.
14. Hazes, B., Magnus, K. A., Bonaventura, C., Bonaventura, J., Dauter, Z., Kalk, K. H. & Hol, W. G. J. (1993) *Protein Sci.* **2,** 597.
15. Teplyakov, A., Obmolova, G., Wilson, K. & Korimuzu, K. (1993) *Eur. J. Biochem.* **213,** 737.
16. Martinez, S. E., Huang, D., Szczepaniak, A., Cramer, W. & Smith, J. L. (1994) *Structure* **2,** 95–105.
17. Jones, E. Y., Davis, S. J., Williams, A. F., Harlos, K. & Stuart, D. I. (1992) *Nature (London)* **360,** 232–239.
18. Bodian, D. L., Jones, E. Y., Harlos, K., Stuart, D. I. & Davis, S. J. (1994) *Structure* **2,** 755–766.
19. Jones, E. Y., Harlos, K., Bottomley, M. J., Robinson, R. C., Driscoll, P. C., Edwards, R. M., Clements, J. M., Dudgeon, T. J. & Stuart, D. I. (1995) *Nature (London)* **373,** 539–544.
20. Ghosh, G., Van Duyne, G., Ghosh, S. & Sigler, P. B. (1995) *Nature (London)* **373,** 303–310.
21. Rossmann, M. G. & Argos, P. (1975) *J. Biol. Chem.* **250,** 3910–3921.
22. Hendrickson, W. A. (1979) *Acta Crystallogr. A* **35,** 158–163.
23. Harpaz, Y. & Chothia, C. (1994) *J. Mol. Biol.* **238,** 528–539.
24. Richardson, J. S. (1981) *Adv. Protein Chem.* **34,** 167–339.
25. Richardson, J. S., Richardson, D. C. & Thomas, K. A. (1976) *J. Mol. Biol.* **102,** 221–235.
26. Hazes, B. & Hol, W. G. (1992) *Proteins* **12,** 278–298.
27. Holmgren, A. & Brändén, C.-I. (1989) *Nature (London)* **342,** 248.
28. Juy, M., Amit, A. G., Alzari, P. M., Poljak, R. J., Claeyssens, M., Beguin, P. & Aubert, J.-P. (1992) *Nature (London)* **357,** 89–91.
29. Cho, Y., Gorina, S., Jeffrey, P. D. & Pavletich, N. P. (1994) *Science* **265,** 346–355.
30. Müller, C. W., Rey, F. A., Sodeoka, M., Verdine, G. L. & Harrison, S. C. (1995) *Nature (London)* **373,** 311–317.
31. Wagner, G. (1995) *Science* **267,** 342.
32. Reid, R. A. & Hemperly, J. J. (1990) *Nucleic Acids Res.* **18,** 5896.
33. Aronson, H.-E. G., Royer, W. E., Jr. & Hendrickson, W. A. (1994) *Protein Sci.* **3,** 1706–1711.
34. Patthy, L. (1991) *Curr. Opin. Struct. Biol.* **1,** 351–361.
35. Puttagunta, S., Mathur, M. & Cowin, P. (1994) *J. Biol. Chem.* **269,** 1949–1955.
36. Brändén, C. & Tooze, J. (1991) *Introduction to Protein Structure* (Garland, New York).
37. Leahy, D. J., Hendrickson, W. A., Aukhil, I. & Erickson, H. (1992) *Science* **258,** 987–991.