

C-Cadherin Ectodomain Structure and Implications for Cell Adhesion Mechanisms

Titus J. Boggon,^{1,4} John Murray,^{1,4} Sophie Chappuis-Flament,⁵ Ellen Wong,⁵ Barry M. Gumbiner,⁵ Lawrence Shapiro^{1,2,3,4*}

Cadherins are transmembrane proteins that mediate adhesion between cells in the solid tissues of animals. Here we present the 3.1 angstrom resolution crystal structure of the whole, functional extracellular domain from C-cadherin, a representative "classical" cadherin. The structure suggests a molecular mechanism for adhesion between cells by classical cadherins, and it provides a new framework for understanding both cis (same cell) and trans (juxtaposed cell) cadherin interactions. The trans adhesive interface is a twofold symmetric interaction defined by a conserved tryptophan side chain at the membrane-distal end of a cadherin molecule from one cell, which inserts into a hydrophobic pocket at the membrane-distal end of a cadherin molecule from the opposing cell.

Cadherins are thought to be the primary mediators of adhesion between the cells of vertebrate animals, and they also function in cell adhesion in many invertebrates (1–4). The expression of numerous cadherins during development is highly regulated, and the precise pattern of cadherin expression plays a pivotal role in the morphogenesis of tissues and organs (1). Furthermore, cadherins are important in the continued maintenance of tissue structure and integrity; for example, loss of cadherin expression appears to be highly correlated with the invasiveness of some types of tumors (5). Cell adhesion mediated by cadherins is thought to be homotypic (6). That is to say, a cell expressing type X cadherin will associate with another cell expressing cadherin X. Cadherin adhesion is also dependent on the presence of millimolar calcium ion concentrations (7), as are found in the extracellular milieu.

The cadherin protein superfamily, defined as proteins containing a cadherin-like domain, can be divided into several subgroups (8). These include the classical (type I) cadherins, which mediate adhesion at adherens junctions; the highly related type II cadherins; the desmosomal cadherins found in desmosome junctions; protocadherins, expressed primarily in the nervous system; and atypical cadherin-like domain containing proteins. Members of all but

the atypical group have been shown to play a role in intercellular adhesion.

Classical, type II, and desmosomal cadherins share a common domain organization: Each comprises five tandem extracellular cadherin domains, a single transmembrane seg-

ment, and a highly conserved cytoplasmic domain (2). Several intermolecular interfaces have been identified in cadherin crystal structures (9, 10), such as the "strand dimer," an interface comprised of a conserved tryptophan side chain (from Trp²) that intercalates into a conserved hydrophobic pocket in a partner molecule. Here, we show that, in the crystal structure of the C-cadherin ectodomain, the strand dimer appears in an orientation poised for adhesion between cadherins presented from adjacent cells. The dependence of cadherin-based cell adhesion on the fidelity of the elements of this interface suggests that the strand dimer is critical for the adhesive function of the classical cadherins. Although the functional data cannot definitively distinguish between a role for this interface in cis or trans interactions, our structure suggests that it may directly participate in the adhesive interaction. Furthermore, the simple twofold symmetry of this interface suggests a rationale for the homophilic specificity that is generally observed in cadherins and reveals the molecular regions of the likely determinants of cadherin specificity.

The structure of the C-cadherin ectodomain was determined to 3.08 Å resolution (11) (Table 1). The overall structure, composed of five

Table 1. Statistics from the crystallographic analysis. The native data set was collected on beamline X4A of the NSLS with a Quantum 4 charge-coupled device detector. $R_{\text{sym}} = \sum |I - \langle I \rangle| / \sum I$, where I is observed intensity and $\langle I \rangle$ is average intensity. $R_{\text{cryst}} = 100 \times \sum ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum |F_{\text{obs}}|$ where F_{obs} are the observed structure factors and F_{calc} are the calculated structure factors. Mean $\langle I/\sigma \rangle$ is above three in all orthogonal directions at 3.4 Å resolution [measured using the program TRUNCATE (37)]. The crystallographic R factor, R_{cryst} , is based on 95% of the data used in refinement, and the free R factor, R_{free} , is based on 5% of the data withheld for the cross-validation test. Meas., measured; p/w/c/s/, protein/water/calcium/sugar; Rmsd, root mean square deviation. Ramachandran angle data (plots of most favorable regions) were calculated with the program PROCHECK (38).

Data collection			
Diffraction maximum (Å)	3.08		
λ used for data collection (Å)	1.000 (at 12398 eV)		
Number of images used for integration	145		
Oscillation range (°)	1		
Resolution range (Å)	20.0 to 3.08		
Number of reflections (meas./unique)	203026/21748		
Coverage (last bin, 3.19 to 3.08 Å)	98.5 (98.3)		
Overall $\langle I/\sigma \rangle$ (last bin, 3.11 to 3.08 Å)	7.68 (2.06)		
Multiplicity	2.7		
R_{sym} (last bin, 3.19 to 3.08 Å)	0.109 (0.372)		
Space group	C 2		
Unit cell	$a = 127.2 \text{ \AA}$, $b = 75.1 \text{ \AA}$, $c = 129.8 \text{ \AA}$, $\alpha = 90.0^\circ$, $\beta = 105.5^\circ$, $\gamma = 90.0^\circ$		
Refinement			
Resolution range (Å)	20.0 to 3.08		
Sigma cutoff (σ)	3 σ		
Number of working set reflections ($ F > 3\sigma$)	14606		
Number of R_{free} set reflections	735		
R_{cryst} (%)	24.3		
R_{free} (%)	27.6		
Number of atoms (p/w/c/s)	4032/38/12/210		
Rmsd from ideal geometry			
Bonds (Å)	0.015	Angles (°)	3.2
Dihedral angles (°)	26.9	Improper angles (°)	2.43
Ramachandran plot, most favorable (%)	69.6		
Additional allowed (%)	28.2		
Generously allowed (%)	2.2		
Disallowed (%)	0.0		

¹Departments of Biochemistry and Molecular Biophysics, ²Department of Ophthalmology, ³Naomi Berrie Diabetes Center, Columbia University College of Physicians and Surgeons, 630 West 168th Street, New York, NY 10032, USA. ⁴Structural Biology Program, Department of Physiology and Biophysics, Mount Sinai School of Medicine, New York, NY 10029, USA. ⁵Cellular Biochemistry and Biophysics Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10021, USA.

*To whom correspondence should be addressed. E-mail: Shapiro@convex.hhmi.columbia.edu

extracellular cadherin-like domains named EC1 (membrane distal) to EC5 (membrane proximal), is elongated but substantially curved such that the long axis of EC1 is approximately perpendicular to the long axis of EC5 (Fig. 1, A and B). This curve is similar to that seen in electron micrographs of isolated natural and recombinant Ca^{2+} -bound E-cadherin extracellular domains (12, 13). These findings suggest that Ca^{2+} -bound classical cadherins may, in general, adopt a stable curved structure over the length of their extracellular domains.

As expected from sequence analysis, each EC domain adopts a “greek key” topology like those observed in earlier structures of cadherin extracellular domain fragments (fig. S1, table S1, and supporting online text) (9, 10, 14–17). Each interdomain boundary (fig. S2 and table S2) is made rigid by three ligated Ca^{2+} ions, with 12 found in the complete structure (fig. S3). Two disulfide bonds (Cys⁴⁴⁸–Cys⁵³² and Cys⁵³⁰–Cys⁵³⁹ are evident in domain EC5 near the membrane-proximal COOH-terminus of the structure.

In the course of refinement, additional electron density appeared to protrude from some

threonine and asparagine residue side chains but no others, suggesting that this density represented both N-linked (Asn) and O-linked (Thr) sugars (Fig. 1D). The glycosylation sites of classical cadherins have not been mapped chemically; however, the electron density maps presented here identify three N-linked and 12 O-linked glycosylation sites (fig. S4). The glycosylation sites are all in domains EC3 and EC4, except for a single O-linked site in domain EC2 and one N-linked site in EC5. The multiply O-glycosylated sequence motif T-X-T-X-T is found in the G strands of both domains EC3 and EC4, and this sequence motif is conserved among the classical cadherins, which also suggests a potential biological role. However, the function of these conserved glycosylation sites is not yet clear.

The primary function of cadherins is to bond juxtaposing cells together through their binding interactions. C-cadherin mediates this function through homophilic binding, and this function is maintained by the recombinant EC1 to EC5 construct described here (18, 19). Thus, it is possible that molecular interfaces observed in crystals of this molecule will reproduce the in

vivo interfaces. In particular, we present evidence that one of the interfaces described here, the strand dimer, may correspond to the primary molecular interface formed between cadherins presented by opposing cells.

The C-cadherin structure reveals a twofold symmetric exchange of the NH_2 -terminal β strands between the EC1 domains of adjoining molecules (Fig. 2). These molecules are arranged in an antiparallel fashion, as if protruding from opposing cell surfaces. A similar strand dimer interaction was first observed in three crystal structures of the NH_2 -terminal domain from N-cadherin (9) (fig. S5 and supporting online text). It was then interpreted—because the COOH-termini of partner molecules were aligned in similar directions—to represent a cis interaction, dimerizing cadherins in parallel from the same cell surface. In the C-cadherin structure, the complete ectodomains are clearly disposed in opposite directions, as if emanating from opposing cell surfaces. Site-directed mutagenesis experiments demonstrate the functional requirement of strand dimer residues for cell adhesion (16, 17, 20–22). However, these experiments cannot formally distinguish between a functional role in cis or trans interactions. The structure presented here suggests that the strand dimer is likely to function in trans interactions. The sequence conservation of strand dimer elements in all classical cadherins (8) (fig. S4) suggests that the strand dimer likely represents the common adhesive mode of classical cadherins.

The strand dimer interface is defined by an exchange of EC1 domain NH_2 -terminal β strands and the completion of EC1 hydrophobic cores by the symmetrical insertion of the conserved Trp² side chain from the partner molecule. Thus, juxtaposing C-cadherin molecules are bound together through an interaction involving conserved hydrophobic core residues in the EC1 domain. This interface defines a twofold symmetric, or reciprocated, ball-and-socket joint.

The strand dimer interface also includes other interactions mediated by residues conserved in different cadherins. The side chain of Glu⁸⁹, conserved in classical cadherins, makes a salt bridge to the NH_2 -terminus. Notably, mutation of Glu⁸⁹ to Ala abolishes adhesion in R-cadherin (21), and extensions of the NH_2 -terminus can also eliminate adhesive function (23). The ring Ne of Trp² forms a hydrogen bond to the backbone carbonyl of residue 90, and residues 1 to 3 of the A strand form antiparallel β conformation hydrogen bonds with residues 27 to 25 of the partner B strand (table S3). Classical, type II, and desmosomal cadherins contain a predomain between the signal sequence and the start of the EC1 domain (2, 4). This predomain must be cleaved to activate the adhesive function (23). The position of the small predomain (usually fewer than 80 amino acids) at the NH_2 -terminus suggests a likely mode for its

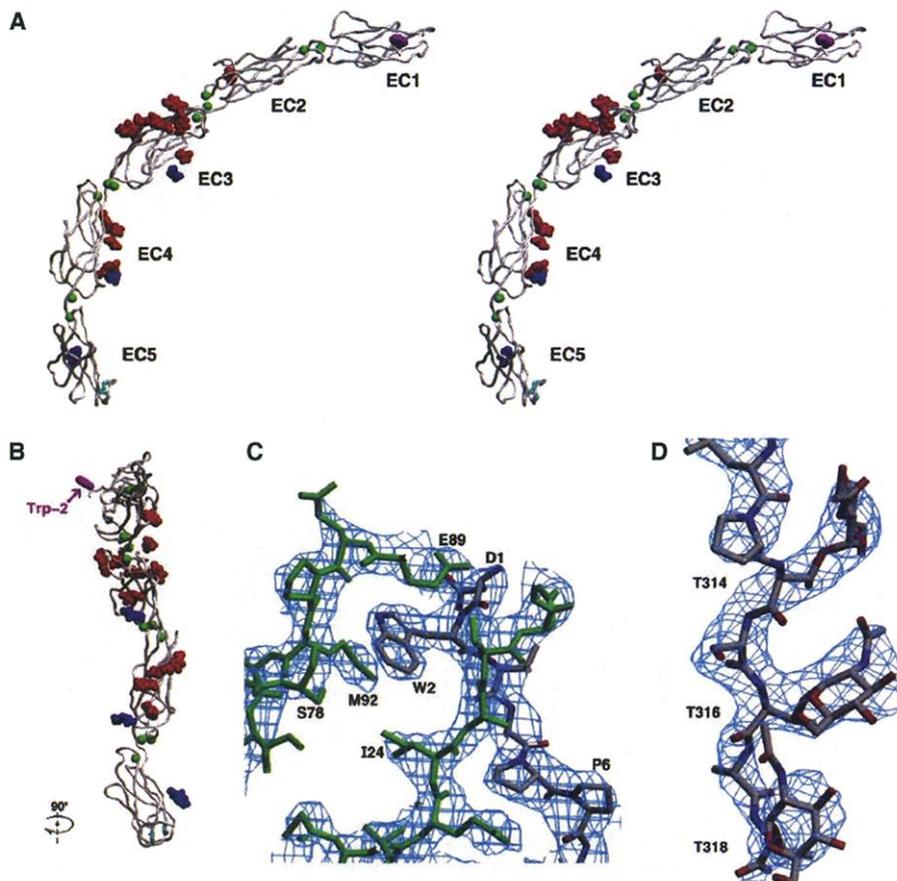


Fig. 1. (A) Stereo view of the C-cadherin ectodomain. Trp² is shown in CPK representation and colored purple; green spheres, calcium ions; cyan, disulfide bonds; red, O-linked sugars; blue, N-linked sugars. Individual cadherin-like domains are labeled EC1 through EC5. (B) View 90° away from (A). (C) An example region of the $3f_{\text{obs}} - 2f_{\text{calc}}$ electron density map contoured at 2σ centered on the strand dimer interface. (D) An example of the $2\sigma 3f_{\text{obs}} - 2f_{\text{calc}}$ density seen for glycosylation sites. Images made with the program SETOR (35).

function through direct occlusion of the adhesive interface.

The C-cadherin structure reveals another interaction that may have biological relevance. This interface (Fig. 3) is formed between the Trp² distal face of an EC1 domain and the bottom of domain 2 (near the COOH-terminal exit point) and is arranged in a parallel fashion, as if to form cis interactions between molecules emanating from the same cell surface. Many experimental results suggest that cis interactions may be an important feature of cadherin adhesive function, likely to promote adhesion through increasing avid-

ity via molecular clustering (13, 16, 18, 20, 24–27).

The cis-oriented interface described here depicts protomers set in a front-to-back arrangement, such that a continuous line of molecules is formed rather than a discrete dimer. This interface involves a complementary fit between the convex surface presented at the bottom of domain 2 (lower parts of the B, D, and E strands) and the concave surface of the EC1 domain (strands C, D, and F, capped at the top by the FG loop and at the bottom by the quasi β helix). Though a number of residues important to this interface are conserved among dif-

ferent classical cadherins, there are also substantial differences (fig. S4). The quasi β helix, which only occurs in the EC1 domains of type I and desmosomal cadherins (8), contains conserved acidic residues, potentially capable of coordinating Ca^{2+} . In the cis-oriented interface reported here, the side chain of Asp⁴⁴ from the quasi β helix is nearly poised to coordinate Ca1 of the EC2-EC3 linker region. Mutation of Asp⁴⁴ to Ala abolishes the adhesive capacity of R-cadherin (21).

Several factors suggest the potential biological relevance of the cis-oriented interaction reported here. First, this interface is es-

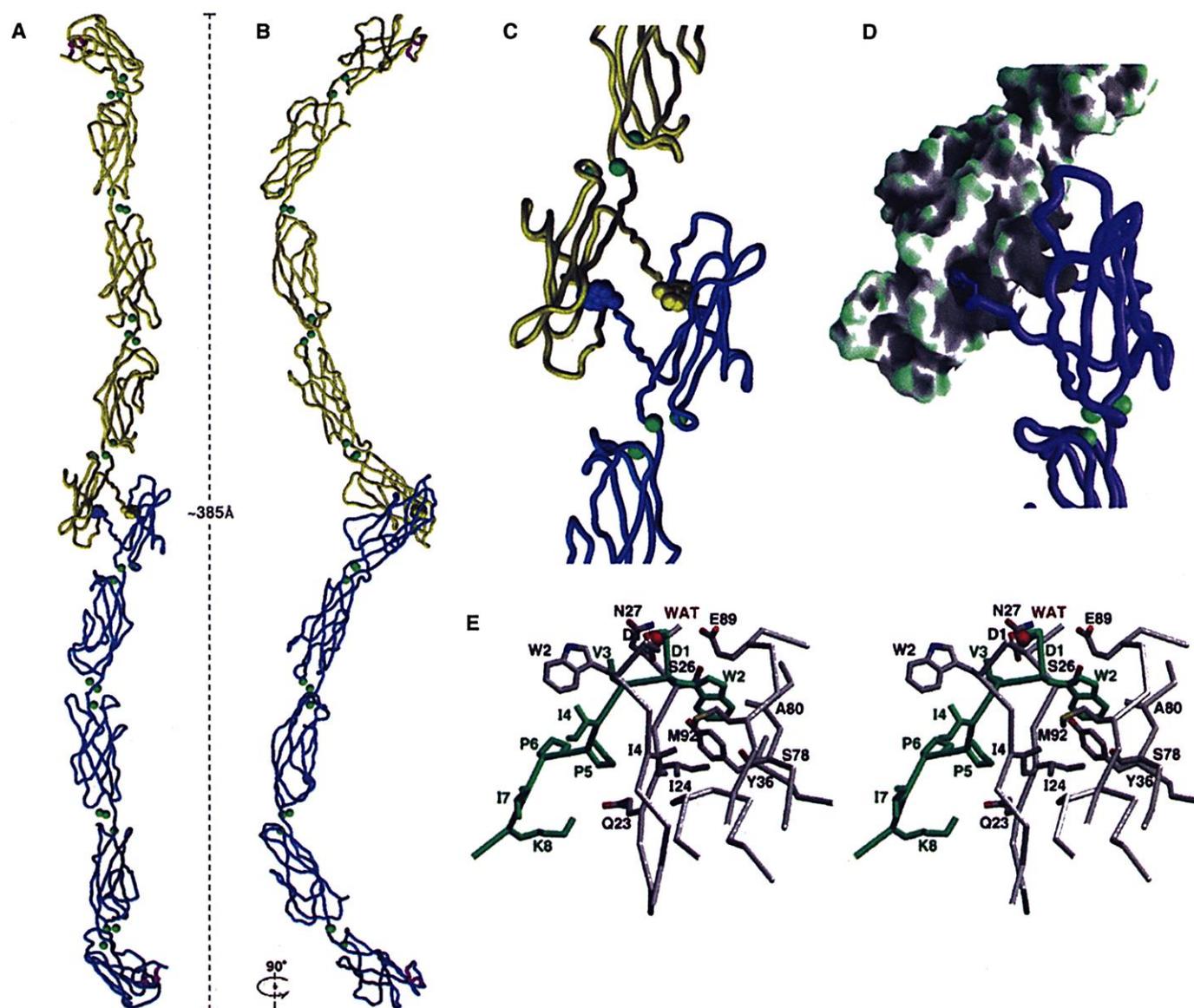


Fig. 2. Strand dimer interface. (A) Experimentally derived backbone worm trace of two C-cadherin ectodomains connected through a strand dimer interaction. The molecules are oriented as if emanating from opposing cell surfaces. Trp² side chains are shown in CPK representation; green spheres, calcium ions; purple, disulfide bonds. (B) View from 90° away from (A). (C) An expanded view of the partner EC1 domains. (D) The strand dimer interface derives from the interaction of highly complementary cross-intercalating surfaces. A molecular surface is shown for one partner molecule, with the other drawn as a backbone worm. Convex surface features are

drawn in green and concave features are gray. Trp² side chains can be seen inserting into a large concave cavity of the opposing EC1 domain. (E) Detailed stereo view of the strand dimer interaction. Side chains that make direct contact with the partner molecule are labeled, and a water molecule that mediates the Asp¹-Asp¹ interaction is shown. (A), (B), (C), and (E) were made with the program SETOR (35), and (D) was made with Grasp (36). Comparisons between the C-cadherin strand dimer interface presented here and previously observed strand dimer interfaces can be found in figs. S5 and S6 and tables S3 and S4.

essentially identical to one present, but not commented on, in two different structures of two-domain fragments from E-cadherin (10, 17). [However, this is not the cis dimerization interface proposed previously for E-cadherin (10) (supporting online text).] Second, recent work has shown that cadherins truncated in the COOH-terminal portion of their extracellular segments must include three extracellular domains (EC1 to EC3) in order to achieve full adhesive capacity (19). Notably, three domains must be present to preserve the EC2 to EC3 Ca²⁺ binding region, which is implicated as potentially important to the cis-oriented interface reported here. Lastly, some experiments suggest that cis dimerization is dependent on high levels of Ca²⁺ (7, 18). The structure of this interface, through its in-

volvement of the EC2-EC3 calcium binding site, provides a plausible explanation for this observation. Electron micrographs of both desmosomes and zonula adherens are suggestive of a protein lattice that might function in maintaining the very uniform intermembrane spacings of these junctions (28, 29). The combination of cis and trans interactions reported here would form a supramolecular complex in agreement with the geometrical constraints of junction formation. The lattice, corresponding to a layer of the C-cadherin crystal (Fig. 4), can provide only a model for a potential mode of cadherin function in the absence of further evidence.

The crystal structure of the extracellular domain from C-cadherin shows the possibility that the strand dimer represents the primary

adhesive interface for classical and desmosomal cadherins. This possibility is supported by existing biochemical data on cadherins: (i) Trp² is conserved in classical and desmosomal cadherins, as are the small hydrophobic residues that line the pocket in which it inserts (the strand dimer "acceptor pocket") (8). (ii) Mutation of either Trp² or residues in the acceptor pocket abolish the adhesive function of classical cadherins (16, 17, 20–22). (iii) Additions of even a few amino acids at the NH₂-terminus can abolish adhesive function (23), suggesting an important functional role at or near the NH₂-terminus. These findings cannot distinguish between cis and trans dimerization, but a fourth line of evidence provides a way to distinguish cis and trans interactions: Electron microscopy studies of a recombinant pentameric E-cadherin fusion protein show that mutation of Trp² to Ala and the pocket mutation Ala⁸⁰ to Leu both abolish the formation of trans intermolecular interactions, but not cis interactions (17, 30). From this observation and their crystal structure of a two-domain E-cadherin fragment, these authors proposed a model in which the Trp² acts intramolecularly to activate another binding interface. However, another interpretation is that Trp² directly mediates the intermolecular interaction. As a caveat, however, we note that it cannot be proved now whether trans interactions between recombinant pentamers correspond to the trans interactions functional between opposing cells.

Other models for the molecular structure of the homophilic adhesive contact between cadherins have been proposed. The prevailing model, the "linear zipper," was based on two interactions observed in the crystal structure of the first NH₂-terminal domain of N-cadherin: a cis interaction resulting from the Trp²-dependent strand dimer interface, and a putative trans adhesion dimer interface between the large surfaces containing the HAV (31) sequence (9). Although this putative adhesive dimer interface involved a large buried surface, site-directed mutagenesis of residues in this interface failed to show substantial effects in the aggregation properties of transfected cells (W.-S. Shan, D. R. Colman, and L. Shapiro, unpublished results). Thus, we now believe that this interface may have resulted from nonphysiological crystal packing interactions. The interactions observed in the present crystal structure of C-cadherin are more likely to be physiologically relevant because they arise in the context of the whole ectodomain rather than a single-domain fragment, and they are in good agreement with the mutagenesis data currently available in the literature.

Alternative models implicating multiple cadherin domains in forming the adhesive interface have been derived from functional analysis of C-cadherin ectodomain deletion constructs (19) and the measurement of unbinding trajectories for the C-cadherin

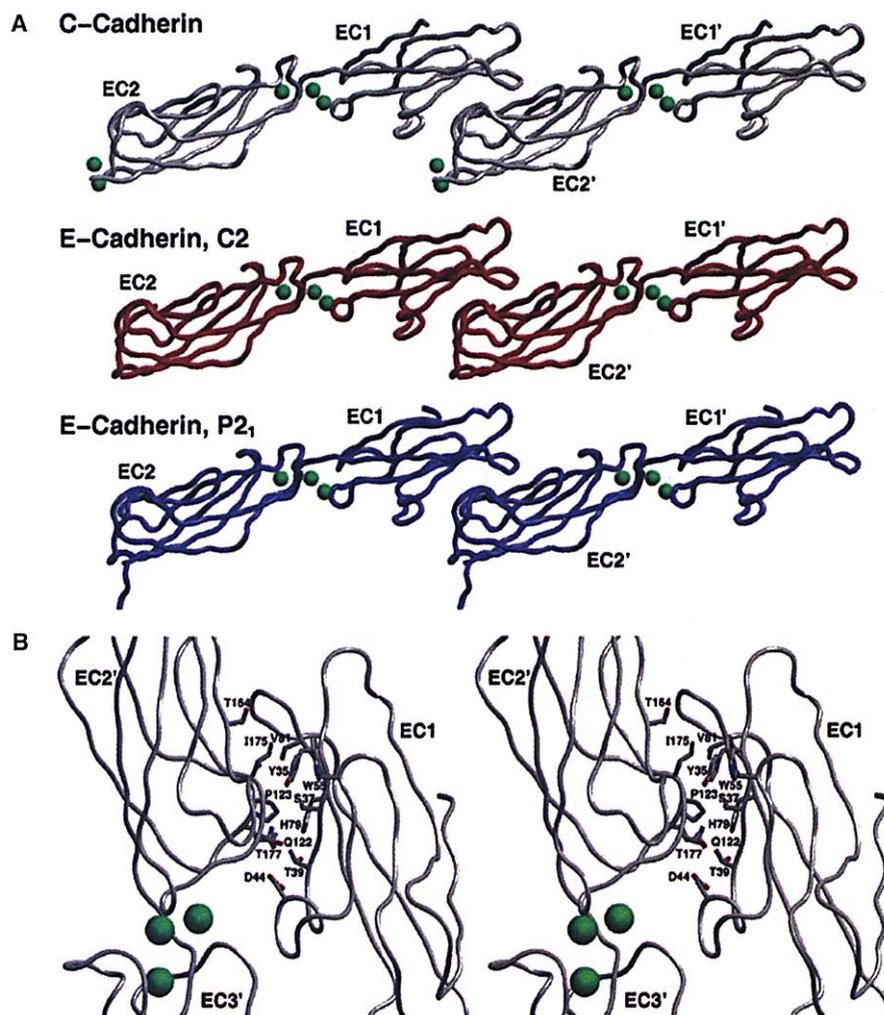
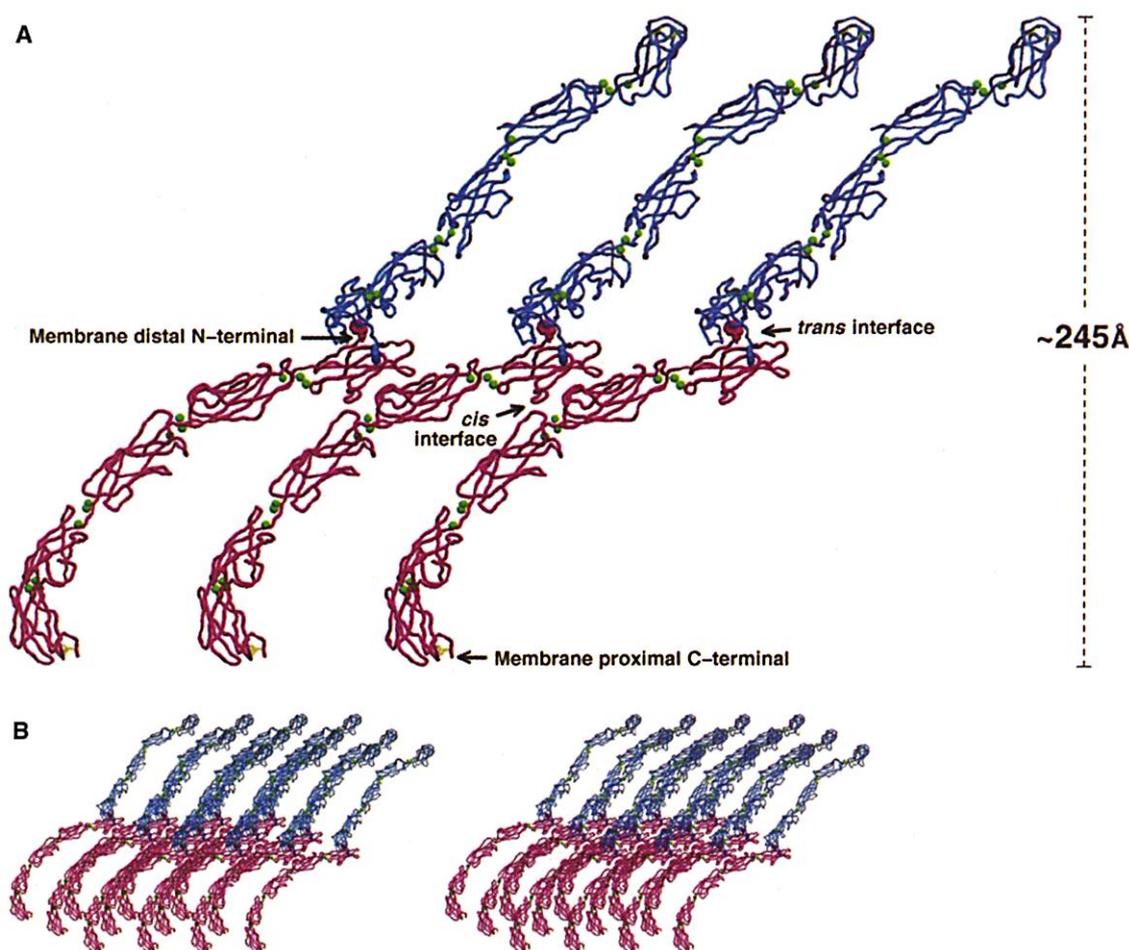


Fig. 3. Potential cis dimerization interface. (A) Comparison of similar interfaces seen in the crystal structures of the C-cadherin ectodomain (this work), and two structures of a two-domain fragment from E-cadherin in space groups P2₁ (PDB code 1FF5) and C2 (PDB code 1EDH). EC1 and EC2 domains are shown. Primary regions of contact include the front β sheet of EC1 (strands C, D, and F), which interact with residues mainly from the back sheet of EC2 (strands B, D, and E). (B) Detailed stereo view of the cis-oriented interface. Residues that make direct contact in the interface are labeled, and calcium ions are shown as green spheres. Primes indicate the partner molecule. The conserved residue Asp⁴⁴ in the quasi- β -helix region is in close proximity to the EC2 to EC3 calcium-binding region, although direct ligation of calcium ions is not observed. Images were made with the program SETOR (35).

Fig. 4. Combination of the cis and trans interfaces results in a lattice, corresponding to one layer of the C-cadherin crystals, which may provide a model for classical cadherin function in cell adhesion. The images in this figure are experimentally determined, meaning they have been observed in the crystal lattice. **(A)** C-cadherin ectodomains, joined by both cis and trans interfaces, are arrayed as if emanating from juxtaposed cell surfaces. Molecules from either putative cell surface are shown in blue or pink. Trp² side chains are shown in CPK representation; green spheres, calcium ions; yellow, disulfide bonds. **(B)** Stereo view of the three-dimensional lattice. Images were made with the program SETOR (35). Comparisons of previously observed cis and trans interfaces are shown in table S5.



ectodomain by surface force measurements (supporting online text) (32). Functional analysis of C-cadherin deletion constructs shows that a molecule with a truncated extracellular domain including only EC1 and EC2 has residual adhesive activity. Addition of EC3, to make an EC1 to EC3 extracellular domain, or EC4 to make an EC1-EC2-EC4 construct, restores full activity, but this restoration cannot be achieved through addition of a fibronectin type III (FNIII) domain forming an EC1-EC2-FNIII chimera (19). These observations can be interpreted as evidence for a direct role for multiple domains in the adhesive interaction, but they are also consistent with a role for the potential cis interface pointed out in the C-cadherin structure. Further functional studies will be required to distinguish between these models and to more fully understand the molecular basis of cadherin adhesion.

Recent studies have shown that classical cadherins are far more promiscuous in their binding interactions with other classical cadherins than was previously thought (33). The relatively small surface area of the strand dimer interface and the conservation of its core elements in all classical cadherins provide a plausible structural basis for the mo-

lecular binding properties of this protein family. Cell-based experiments demonstrate the function of the strand dimer in cis interactions between cadherin ectodomains (20, 22, 34), whereas electron microscopic studies indicate that strand dimer elements can contribute to trans interactions (17, 30). These data can potentially be reconciled in light of the structural plasticity of the strand dimer, which may enable it to adopt conformations that accommodate either orientation.

Given the lack of sequence specificity of the primary strand dimer interactions, what, then, could provide determinants for cadherin adhesive specificity? The only substantial region of contact between strand dimer-related molecules—aside from the Trp² interface itself—is between the A strand of one protomer and the B strand of its partner (supporting online text). Desmosomal cadherins differ substantially in this region (8), but the classical cadherins are highly conserved with only a few exceptions, and these are poised to contact one another—perhaps providing the basis for the specificity observed among classical cadherins. It is possible that even small changes in molecular affinity will result in substantially different amplified avidities at intercellular junctions containing clusters of

cadherin molecules. Nonetheless, the relative promiscuity observed for interactions between different cadherins agrees well with the strand dimer model.

References and Notes

1. B. M. Gumbiner, *J. Cell Biol.* **148**, 399 (2000).
2. M. Takeichi, *Annu. Rev. Biochem.* **59**, 237 (1990).
3. M. Takeichi, S. Nakagawa, S. Aono, T. Usui, T. Uemura, *Philos. Trans. R. Soc. London B* **355**, 885 (2000).
4. B. Gumbiner, B. Stevenson, A. Grimaldi, *J. Cell Biol.* **107**, 1575 (1988).
5. G. Bex, F. Nollet, F. van Roy, *Cell Adhes. Commun.* **6**, 171 (1998).
6. M. S. Steinberg, P. M. McNutt, *Curr. Opin. Cell Biol.* **11**, 554 (1999).
7. A. W. Koch, S. Pokutta, A. Lustig, J. Engel, *Biochemistry* **36**, 7697 (1997).
8. F. Nollet, P. Kools, F. van Roy, *J. Mol. Biol.* **299**, 551 (2000).
9. L. Shapiro *et al.*, *Nature* **374**, 327 (1995).
10. B. Nagar, M. Overduin, M. Ikura, J. M. Rini, *Nature* **380**, 360 (1996).
11. Materials and methods are available as supporting online material.
12. S. Pokutta, K. Herrenknecht, R. Kemler, J. Engel, *Eur. J. Biochem.* **223**, 1019 (1994).
13. A. Tomschy, C. Fauser, R. Landwehr, J. Engel, *EMBO J.* **15**, 3507 (1996).
14. M. Overduin *et al.*, *Science* **267**, 386 (1995).
15. L. Shapiro, P. D. Kwong, A. M. Fannon, D. R. Colman, W. A. Hendrickson, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6793 (1995).
16. K. Tamura, W. S. Shan, W. A. Hendrickson, D. R. Colman, L. Shapiro, *Neuron* **20**, 1153 (1998).
17. O. Pertz *et al.*, *EMBO J.* **18**, 1738 (1999).

18. W. M. Brieher, A. S. Yap, B. M. Gumbiner, *J. Cell Biol.* **135**, 487 (1996).
19. S. Chappuis-Flament, E. Wong, L. D. Hicks, C. M. Kay, B. M. Gumbiner, *J. Cell Biol.* **154**, 231 (2001).
20. W. S. Shan et al., *J. Cell Biol.* **148**, 579 (2000).
21. M. Kitagawa et al., *Biochem. Biophys. Res. Commun.* **271**, 358 (2000).
22. R. B. Troyanovsky, J. Klingelhofer, S. Troyanovsky, *J. Cell Sci.* **112**, 4379 (1999).
23. M. Ozawa, R. Kemler, *J. Cell Biol.* **111**, 1645 (1990).
24. A. S. Yap, W. M. Brieher, M. Pruschy, B. M. Gumbiner, *Curr. Biol.* **7**, 308 (1997).
25. A. S. Yap, C. M. Niessen, B. M. Gumbiner, *J. Cell Biol.* **141**, 779 (1998).
26. H. Takeda, Y. Shimoyama, A. Nagafuchi, S. Hirohashi, *Nature Struct. Biol.* **6**, 310 (1999).
27. S. M. Troyanovsky, *Curr. Opin. Cell Biol.* **11**, 561 (1999).
28. K. Miyaguchi, *J. Struct. Biol.* **132**, 169 (2000).
29. L. A. Staehelin, *Int. Rev. Cytol.* **39**, 191 (1974).
30. A. W. Koch, D. Bozic, O. Pertz, J. Engel, *Curr. Opin. Struct. Biol.* **9**, 275 (1999).
31. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
32. S. Sivasankar, B. Gumbiner, D. Leckband, *Biophys. J.* **80**, 1758 (2001).
33. C. M. Niessen, B. Gumbiner, *J. Cell Biol.* **156**, 389 (2002).
34. H. Tanihara, K. Sano, R. L. Heimark, T. St. John, S. Suzuki, *Cell Adhes. Commun.* **2**, 15 (1994).
35. S. V. Evans, *J. Mol. Graph.* **11**, 134 (1993).
36. A. Nicholls, K. Sharp, B. Honig, *Proteins* **11**, 281 (1991).
37. Collaborative Computational Project No. 4, *Acta Crystallogr. D* **50**, 760 (1994).
38. R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, *J. Appl. Crystallogr.* **26**, 283 (1993).
39. We are grateful to C. Ogata and the staff of the NSLS beamline X4A for help with data collection. We thank W. A. Hendrickson, P. Kwong, R. Axel, and T. M. Jessell for many helpful discussions. T.J.B. was the recipient of a Wellcome Trust International Prize Travelling Research Fellowship (056509/Z/98/Z). Supported by

an NIH grant (R01 GM52717) awarded to B.M.G.; the Dewitt Wallace Fund for Memorial Sloan-Kettering Cancer Center; a Cancer Support grant (NCI-P30-CA-08784); an NIH grant (R01 GM062270) awarded to L.S., who was also the recipient of a Career Scientist Award from the Irma T. Hirsch Foundation; and a Career Development Award from the American Diabetes Association. Beamline X4A at the NSLS, a DOE facility, is supported by the Howard Hughes Medical Institute. Coordinates have been deposited in the Protein Data Bank (accession code 1L3W).

Supporting Online Material
(www.sciencemag.org/cgi/content/full/1071559/DC1)
Materials and Methods
Supporting Text
figs. S1 through S6
tables S1 through S5

6 March 2002; accepted 9 April 2002
Published online 18 April 2002;
10.1126/science.1071559
Include this information when citing this paper.

Vitamin D Receptor As an Intestinal Bile Acid Sensor

Makoto Makishima,^{1*} Timothy T. Lu,¹ Wen Xie,^{2†}
G. Kerr Whitfield,³ Hideharu Domoto,¹ Ronald M. Evans,²
Mark R. Haussler,³ David J. Mangelsdorf^{1‡}

The vitamin D receptor (VDR) mediates the effects of the calcemic hormone 1 α ,25-dihydroxyvitamin D₃ [1,25(OH)₂D₃]. We show that VDR also functions as a receptor for the secondary bile acid lithocholic acid (LCA), which is hepatotoxic and a potential enteric carcinogen. VDR is an order of magnitude more sensitive to LCA and its metabolites than are other nuclear receptors. Activation of VDR by LCA or vitamin D induced expression *in vivo* of CYP3A, a cytochrome P450 enzyme that detoxifies LCA in the liver and intestine. These studies offer a mechanism that may explain the proposed protective effects of vitamin D and its receptor against colon cancer.

A contributing factor to the deleterious effects of a high-fat diet is an associated increase in the excretion of fecal bile acids (1), the most toxic of which is the secondary bile acid LCA (Fig. 1A). Unlike the primary bile acids, chenodeoxycholic acid (CDCA) and cholic acid (CA), LCA is poorly reabsorbed into enterohepatic circulation and passes into the colon. At high concentrations, LCA induces DNA strand breaks, forms DNA ad-

ducts, and inhibits DNA repair enzymes (1–3). LCA can also promote colon cancer in animals (4), and its concentration is higher than other secondary bile acids in patients with colorectal cancer (5).

In contrast to the positive correlation among high-fat diets, LCA, and colon cancer, dietary intake of vitamin D and calcium is related to a reduced incidence of colorectal cancer (6). Furthermore, vitamin D supplementation inhibits colon carcinogenesis induced by either high-fat diets or intrarectal instillation of LCA (7, 8). One route for LCA elimination is through its catabolism by the enterohepatic cytochrome P450, CYP3A, a putative target gene of vitamin D (9, 10). Expression of CYP3A in the liver is regulated by the nuclear xenobiotic and pregnane X receptor (PXR, also called SXR), which can be activated by high concentrations (≥ 100 μ M) of LCA (11, 12). Primary bile acids (in particular, CDCA and CA) are also ligands for the farnesoid X receptor, FXR (13, 14). However, neither PXR nor FXR responds to vitamin D, and LCA-induced expression of CYP3A is still present in PXR-null animals.

This suggests another LCA-dependent pathway for inducing CYP3A expression (11).

To determine if bile acids could act on the vitamin D receptor (VDR) to induce CYP3A expression, we used a ligand-screening assay based on the ligand-induced interaction of a nuclear receptor with its coactivator (14). The receptor-interacting domain of the coactivator SRC-1 was fused to the DNA binding domain of the yeast transcription factor GAL4, and various nuclear receptors were fused to the transactivation domain of the herpes virus VP16 protein. Expression plasmids for GAL4-SRC-1 and VP16-nuclear receptor were transfected with a GAL4-responsive luciferase reporter plasmid into human embryonic kidney (HEK293) cells and examined for luciferase expression after LCA treatment. LCA (30 μ M) induced a ligand-dependent interaction between VDR and SRC-1 (Fig. 1B). As previously reported (14), LCA also activated FXR. However, no other nuclear receptors were activated by LCA (Fig. 1B) (15), including PXR, which required higher LCA concentrations (≥ 100 μ M) to be activated. To further investigate the ligand specificity of VDR and FXR, we tested various primary, secondary, and conjugated bile acids in this assay (Fig. 1C). We performed these experiments in the presence or absence of the ileal bile acid transporter (IBAT), because hydrophilic bile acids such as CA and conjugated bile acids require transport across cell membranes (16). As expected, treatment of cells with the vitamin D hormone 1,25(OH)₂D₃ activated VDR but not FXR (Fig. 1C). Conversely, the primary bile acids CDCA, CA, and their conjugated metabolites were effective ligands for FXR but not VDR. FXR was also activated by the secondary bile acids, deoxycholic acid, LCA, and their conjugated metabolites (Fig. 1C). However, the only bile acids that activated VDR were LCA and its major metabolites 3-keto-LCA (Fig. 1A), glyco-LCA, and 6-ke-

¹Howard Hughes Medical Institute, Department of Pharmacology, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390–9050, USA. ²Howard Hughes Medical Institute, Gene Expression Laboratory, The Salk Institute for Biological Studies, Post Office Box 85800, San Diego, CA 92186–5800, USA. ³Department of Biochemistry and Molecular Biophysics, College of Medicine, University of Arizona, Tucson, AZ 85724, USA.

*Present address: Department of Organismal Biosystems, Osaka University, Suita, Osaka 565-0871, Japan.

†Present address: Center for Pharmacogenetics and Department of Pharmaceutical Sciences, University of Pittsburgh, Pittsburgh, PA 15213, USA.

‡To whom correspondence should be addressed. E-mail: davo.mango@utsouthwestern.edu