Published in final edited form as:

J Mol Biol. 2007 October 19; 373(2): 401–411.

The crystal structure of human E-cadherin domains 1 and 2, and comparison with other cadherins in the context of adhesion mechanism

Emilio Parisini^{1,2}, Jonathan M.G. Higgins³, Jin-huan Liu¹, Michael B. Brenner³, and Jia-huai Wang^{1,4,5}

¹ Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115

² Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115

⁴ Department of Pediatrics, Harvard Medical School, Boston, Massachusetts 02115

⁵ Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115

³ Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115

Summary

Cell adhesion mediated by type I cadherins involves homophilic “*trans*” interactions that are thought to be brought about by a strand exchange mechanism involving the N-terminal extracellular domain. Here, we present the high-resolution crystal structure of the N-terminal two domains of human E-cadherin. Comparison of this structure with other type I cadherin structures reveals features that are likely to be critical to facilitate dimerization by strand exchange as well as dimer flexibility. We integrate this structural knowledge to provide a model for type I cadherin adhesive interactions. Intra-molecular docking of the conserved N-terminal “adhesion arm” into the acceptor pocket in monomeric E-cadherin appears largely identical to inter-molecular docking of the adhesion arm in adhesive *trans* dimers. A strained conformation of the adhesion arm in the monomer, however, may create an equilibrium between “open” and “closed” forms that primes the cadherin for formation of adhesive interactions, which are then stabilized by additional dimer-specific contacts. By contrast, in type II cadherins, strain in the adhesion arm appears absent and a much larger surface area is involved in *trans* adhesion, which may compensate the activating energy paid for peeling off the intra-molecularly docked arm. It seems that evolution has selected slightly different adhesion mechanisms for type I and type II cadherins.

Keywords

cadherin; cell adhesion; crystal structure; adhesion arm

The cadherins comprise a large family of calcium-dependent adhesion molecules that are differentially and, to a large extent, specifically expressed in different tissues. In all metazoa studied to date, cadherin expression on the surface of cells is critical for cell-cell adhesion and,

consequently, for development and for integrity of solid tissues¹. In mammals, impaired cadherin-mediated adhesion has been associated with tumor progression and metastasis².

Phylogenetic analyses of cadherin protein sequences reveal a number of subfamilies: type I and type II cadherins, desmocollins, desmogleins, protocadherins, flamingo cadherins and a variety of other cadherin-related proteins³. These molecules all contain a variable number of immunoglobulin (Ig) like extracellular cadherin (EC) domains. Based on structures of type I, type II and protocadherins, they are likely to be rigidified by the presence of calcium ions at the domain interfaces^{4,5}. Despite these conserved features, other elements of cadherins are quite diverse. For instance, vertebrate type I and type II cadherins, desmogleins and desmocollins all have five EC domains but have divergent cytoplasmic tails that associate with a variety of proteins to provide functional links to the cytoskeleton³. Interestingly, although many cadherins share a capacity for homophilic binding, the molecular mechanisms used by different subfamilies to accomplish this also appear to vary significantly.

Over the last decade, a number of studies have elucidated structure-function relationships in the cadherin family. Most have been conducted using type I cadherins such as E-, N- and C-cadherin, and a “strand-dimer” model has been established for the homophilic adhesion mechanism of these molecules⁴. First proposed as a *cis* interaction based on a structure of N-cadherin EC1⁶, this model was later re-interpreted and widely accepted as a *trans* interaction after structural analysis of C-cadherin EC1–5⁷. The basic feature of this model is a strand exchange mechanism involving the six highly conserved N-terminal residues (dubbed the “adhesion arm” hereafter) of EC1 and a similarly conserved “acceptor pocket” in the body of EC1. To form adhesive contacts, the adhesion arm of a cadherin molecule on the cell surface interacts with the acceptor pocket in EC1 of another molecule on the opposing cell surface. Type II cadherins undergo a similar but distinct strand exchange⁸. Based on sequence analysis and electron tomography of desmosomes⁹, desmocollins and desmogleins are likely to use a similar mechanism. Protocadherins and other cadherin-related molecules, however, appear to use different and still largely unknown mechanisms. The structural factors that bring about the precise balance of binding energies necessary to facilitate strand exchanges in cadherins and the consequences of the differing mechanisms used by different family members are unclear, and may have a direct impact on the adhesion specificity of these molecules.

Here we report the 2.0 Å crystal structure of the N-terminal two EC domains of human E-cadherin. The high resolution of this structure allows us to discuss the “strand-dimer” adhesion mechanism of type I cadherins; to compare it with the dimer interactions of other cadherin family members, in particular type II cadherins; and to highlight the basic principles that are likely to underlie the homophilic binding capacity of these cadherins.

Results and discussion

Figure 1 is a ribbon drawing of the high-resolution human E-cadherin EC1–EC2 structure. The overall topology of the molecule is very similar to that of other cadherin structures (see review⁴). The Ig-like fold, the signature calcium-binding motif at the inter-domain junctions and hence the relatively rigid inter-domain connections are all present. Indeed, the EC1 domain of human E-cadherin EC1–EC2 can be superimposed on murine counterpart (PDB code 1EDH)¹⁰ with a RMSD value of about 0.5 Å, if the first dozen residues are excluded.

Factors that favor inter-molecular adhesion over intra-molecular interaction

The key feature of strand exchange dimerization is that the sidechain of a conserved Trp2 on the adhesion arm inserts into the hydrophobic acceptor pocket of a dyad-related molecule. Comparison of our human E-cadherin EC1–EC2 structure with other representative cadherin structures, the N-cadherin EC1⁶ and C-cadherin EC1–EC5⁷ reveals the following additional

conserved interactions from *trans* adhesive molecule A to molecule B (and vice versa): i) a salt bridge between the N-terminal $^A\text{NH}_3^+$ and the carboxyl group of $^B\text{Glu89}$; ii) mainchain hydrogen bonds between $^A\text{Val3-NH}$ and $^B\text{Lys25-CO}$ as well as between $^A\text{Asp1-CO}$ and $^B\text{Asn27-NH}$; iii) a hydrogen bond between N ϵ of $^A\text{Trp2}$ and carbonyl group of $^B\text{Asp90-CO}$. These forces have been extensively discussed before. In addition we observed another interaction associated with Trp2 that exists in all relevant type I and type II cadherin structures, described below.

The side chain of tryptophan is an indole, comprising a five-membered pyrrole ring fused with a six-membered benzene ring. Figure 2 depicts the sandwich of the planar carboxyl group of $^B\text{Glu89}$ and the aliphatic side chain of $^B\text{Met92}$ from the acceptor pocket onto the pyrrole ring of $^A\text{Trp2}$. The slightly polar nature of the N ϵ -H-group-containing pyrrole ring is stabilized by the negatively charged $^B\text{Glu89}$ and also slightly negatively charged side chain of $^B\text{Met92}$. This stacking configuration of a carboxyl group from a Glu onto the five-membered ring of Trp2 is also found in all other relevant type I and II cadherin structures. This is reminiscent of the cation- π interaction seen in many protein structures between the six-membered ring of a Trp and a positively charged residue such as Arg or Lys ¹¹. Along with the hydrogen bond between the N ϵ -H of $^A\text{Trp2}$ and the mainchain carbonyl group of $^B\text{Asp90}$, this arrangement probably provides an important force to hold the indole group in a fixed position. The interactions between the benzene ring of the indole and other hydrophobic residues in the acceptor pocket further stabilize $^A\text{Trp2}$ anchoring.

Cadherins are synthesized as inactive precursors with a pro-sequence of more than 100 amino acids at the N-terminus, and become active only after proteolytic removal of this domain ¹². The pro-sequence may interfere with dimerization by steric hindrance ¹³, and it also prevents formation of the salt bridge between the mature N-terminal NH_3^+ and the carboxyl group of Glu89 that optimizes docking of Trp2. NMR studies indicate that, at high protein concentration, removal of the pro-sequence favors *trans* dimerization by strand exchange while, at low protein concentration, Trp2 can dock intra-molecularly after the pro-sequence is removed ¹⁴. Indeed, there are two E-cadherin structures in which Trp2 docks into the intra-molecular acceptor pocket: mouse E-cadherin EC1-EC2 (PDB code 1FF5) ¹⁵ and human E-cadherin EC1 in complex with internalin from the bacterium *Listeria* (PDB code 1O6S) ¹⁶. Both these cadherin molecules have “extra” residues N-terminal to the native terminus, precluding formation of a true intra-molecular salt bridge between Asp1- NH_3^+ and Glu89. Nevertheless, it is possible that in the human E-cadherin structure 1O6S, it is the intimate wrapping of internalin around the EC1 monomer that may stabilize the intra-molecularly docked conformation of the adhesion arm. In addition, it is interesting to notice that a bidentate hydrogen bond is found between the amide group of Ser1-NH and Glu89 in this structure (note that Ser replaces the natural Asp1 in this molecule). This hydrogen bond seems to be in the place of the intra-molecular salt bridge that would be expected to form in the native structure. In the case of the mouse E-cadherin structure 1FF5, the single extra N-terminal methionine seems to be insufficient to completely prevent intra-molecular docking, and it has been proposed that crystal packing forces stabilize the normally weak docking interaction in this instance ¹⁴. Overall, these two structures may serve as surrogates for a conformation of native monomeric cadherin, or a form before adhesion. Both clearly show that the interactions between the adhesion arm and EC1 body observed in *trans* adhesive molecules described above can also exist intra-molecularly. We have superimposed the structure of mouse E-cadherin EC1 (1FF5) ¹⁵ with intra-molecular docking of the adhesion arm onto the structure of human E-cadherin EC1 engaged in *trans* dimerization reported here. Figure 3A shows how similar the Trp2 binding in the acceptor pocket is both intra-molecularly (designated the “closed form”) and inter-molecularly (designated the “adhesive form”). This structural observation implies that the free energies of the closed and adhesive forms of the cadherin molecule are quite similar, but it does not explain how the entropic cost of dimerization is overcome.

It is believed that cell adhesion is driven by high avidity and the interaction between individual adhesive molecular pairs is often weak¹⁷. The K_D value for homophilic adhesion of type I cadherins is at the sub-millimolar level¹⁴, and the binding strength of type II cadherins is likely to be comparable^{18; 19}. In fact, the association of the Trp2 is remarkably brief, in the 2 second range²⁰. The expected dynamic equilibrium between the closed and adhesive forms has been elegantly demonstrated in a recent mutagenesis study on N-cadherin focusing on the role of the salt bridge between ^AAsp1-NH₃⁺ and ^BGlu89²¹. Within a wild-type strand-dimer, two salt bridges can form: between ^AAsp1-NH₃⁺ of one molecule and the carboxyl group of ^BGlu89 in the opposing molecule, and vice versa. As expected, a Glu89Ala mutation or adding Gly-Gly to the N-terminus prevents the salt bridge from forming, which impairs adhesion, consistent with our previous results obtained for the Glu89Ala mutation²².

Interestingly, when the two modifications were present separately (with molecule A being a Glu89Ala mutant and molecule B having Gly-Gly added at the N-terminus), stronger adhesion than wild-type was observed. It seems that neither mutant molecule A nor B allows optimal Trp2 docking into its own acceptor pocket due to the loss of the intra-molecular salt bridge. However, the N-terminal ^AAsp1-NH₃⁺ of molecule A can still form a salt bridge to the ^BGlu89 of molecule B, bringing its adhesion arm into place. Since the intra-molecular docking is unfavorable, the otherwise existing activation barrier for the adhesion arm switching from intra-molecular docking to inter-molecular docking may be substantially lowered, leading to enhanced adhesion.

These results imply that for a monomeric cadherin molecule, there is an equilibrium between the closed form with the adhesion arm affixed and an “open” form with the adhesion arm stretched out²¹. Indeed, in a number of structures of E- and N-cadherins, with both native and extended N-termini, the N-terminal part of the adhesion arm is unresolved^{10; 15; 23}, indicating that the arm has a tendency to dissociate from the body of EC1. When an opposing cadherin molecule approaches, *trans* strand-dimer formation is favored, suggesting that the inter-molecular docking of the adhesion arm has a lower free energy than the intra-molecular docking. It is possible that this may be due to in part the existence of two consecutive prolines at the base of the N-terminal segment that introduce strain into the intra-molecularly docked adhesion arm. It has been noticed that the first seven residues (Asp-Trp-Val-Ile-Pro-Pro-Ile) of the N-terminal adhesion arm of type I cadherins are conserved³. In the adhesive form of a cadherin, exemplified by our human E-cadherin EC1-EC2 structure, mainchain hydrogen bonding to form a β -sheet between β -strand A and β -strand G in EC1 occurs only C-terminal to Ile7. N-terminal to Ile7, the Pro5-Pro6 motif ushers the first six residues out away from the body of EC1 (Figure 3B, green molecule). Intriguingly, despite the fact that the Trp2 docks into its own acceptor pocket in the two closed forms (Figure 3B, mouse E-cadherin 1FF5 in yellow and human E-cadherin 1O6S in cyan), the Pro5-Pro6 motif still bulges out prominently. The very N-terminal residues of human E-cadherin 1O6S (cyan) form an anti-parallel pairing with the β strand B (one main chain H-bond between Ser1-CO and Asn27-NH as well as two H-bonds between Val3 and Lys25). Notably, Pro5 in this molecule appears in a different conformation from that in the adhesive form. Its conformational angle Φ is -86° , as opposed to energetically favorable $-55^\circ - -75^\circ$ ²⁴. This proline seems to be in a strained state. In contrast, in the closed structure of mouse E-cadherin 1FF5¹⁵, Pro5-Pro6 can be overlaid more closely on the counterpart in the adhesive form (Figure 3B; yellow molecule overlaid on green). The conformational angle Φ of Pro5 is -58° for the adhesive form reported here (green), within the favorable range, and -77° for the mouse closed form 1FF5 (yellow), not far from the favorable range. In this closed structure, which has an extra N-terminal methionine, Asp1 turns away from forming any interaction with Glu89 (Figure 3B). Although Trp2 is still held in the acceptor pocket, it assumes a very strained conformation ($\Phi = -124$ and $\Psi = 2$). In this case, the strain within the closed form, perhaps introduced by Pro5-Pro6, is manifest at a different position but again it is likely to promote dissociation of the adhesion arm.

Proline-induced arm exchange between protomers plays a key role in protein quaternary structures²⁵. Proline residues found frequently at hinge regions impose constraints on polypeptide conformation and dynamics, and consequently favor extended conformations of arms and oligomerization. Phylogenetic analysis and site-directed mutagenesis experiments all confirm the importance of the existence of proline at these hinge positions^{25; 26}. We consider the Pro5-Pro6 motif in type I cadherins to be a similar structural element acting in a different biological context.

It is tempting to postulate that once the pro-sequence is removed, a type I cadherin molecule is in a meta-stable state, primed for adhesion, in which the adhesion arm is in an equilibrium between the open and closed conformations. When human E-cadherin molecules approach one another from opposing cell surfaces, adhesive *trans* interactions form (Figure 5). During this process, the intra-molecular H-bonds between the N-terminus of the A strand and the B strand will be replaced by similar intermolecular interactions, such as between ^AVal3-NH and ^BLys25-CO. The critical salt bridge between the N-terminal ^AAsp1-NH₃⁺ and ^BGlu89 of the opposing molecule is formed. The indole ring of ^ATrp2 flips from being inwards to outwards, anchoring into the inter-molecular acceptor pocket. Importantly, any strain caused by the Pro5-Pro6 motif will be resolved, resulting in a more stable conformation of the adhesion arm. Intermolecular docking is likely to be further stabilized by the formation of extra contacts upon dimerization. In the dimer, the ^AVal3/^BLys25, ^APro5/^BLys25 and ^APro5/^BGln23 interactions form in an interdigitating fashion (Figure 3C), whereas in the closed internalin complex structure 1O6S there are only two such contacts, ^APro5/^AVal22 and ^AVal3/^ALys25. In addition, our high-resolution structure reveals the presence of two water molecules that appear to stabilize the outward facing conformation of the arm by forming hydrogen bonds with backbone atoms of ^AIle4 (in the arm) and of ^AGlu93 and ^ALeu95 (in the body of the cadherin; Figure 3D). Overall, formation of adhesive interactions is an energetically more favorable process.

Dynamics of the strand-dimer

Figure 4A illustrates a representative collection of EC1 domains from type I cadherin structures overlaid on one another, including human E-cadherin reported here (green), mouse E-cadherin in closed form (PDB code 1FF5; yellow), mouse E-cadherin in adhesive form (PDB code 1Q1P; red) and *Xenopus* C-cadherin (adhesive; PDB code 1L3W; light grey). Other than the closed form in yellow the rest of the structures form strand-dimers. Clearly, the EC1 domains of all the structures superimpose extremely well, with the exception of the first six residues. A remarkable feature here is the wide spectrum of swing angles observed for these six-residue N-terminal segments, and the consequently variable orientations in which the corresponding strand-dimer-associated molecules are found. Crystal packing may have an effect on these differences, and some may be due to intrinsic differences between the type I cadherins⁸. As shown in Figure 4B, however, where just the two strand-dimers of E-cadherin are shown for clarity (human in green/magenta and mouse in red/yellow with the green and red molecules overlaid on top of each other), the orientations of the two dyad-related molecules (yellow and magenta) differ by about 49°. This difference is likely to reflect the dynamic nature of the first six residues in E-cadherin, and defines a characteristic of the type I cadherin family.

As mentioned earlier, several new contacts formed in the E-cadherin dimer appear to facilitate dimerization. It is interesting to note, however, that while these interactions remain mainly hydrophobic in nature, the specific contacts vary with the swing angle of the adhesion arm. For example, in the mouse E-cadherin dimer, the adhesion arms are much more extended than in the human E-cadherin structure, and the interdigitating interactions between ^AVal3 and ^APro5 with ^BGln23 and ^BLys25 mentioned earlier do not form. In this case ^AIle4, which forms intra-molecular hydrophobic contacts with residues including ^AVal22 and ^AMet92 in

the closed structures (Figure 4C) and in the human E-cadherin dimer (Figure 4D), instead forms a close contact with ^BIle4 of the interacting molecule (Figure 4E). This may provide a stabilizing inter-molecular contact that compensates for those weakened upon transition from the less extended conformation seen in the human E-cadherin dimer.

The relative orientations of the E-cadherin molecules in the mouse and human E-cadherin dimers may lie close to the two extremes of a dynamic process described by a shallow free energy curve. Local minima in the energy landscape may explain the ability to crystallize the “intermediates” that are represented by the various structures shown in Figure 4. Indeed, our structure provides some crystallographic data to support the type of conformational changes predicted by molecular dynamic simulations ²⁷. While these conformational differences may be stabilized by crystal packing contacts, we propose that they also reflect real changes in cadherin contacts that occur in response to mechanical forces at cell-cell junctions.

“Cis” interactions

In addition to strand exchange, a second interaction has been observed between E- and C-cadherin molecules in a variety of crystal structures. This so-called “cis” interaction occurs between cadherins in a parallel orientation, as if they are both emanating from the same cell surface. In this case, a hydrophobic surface on the “back” of domain 1, away from the site of strand exchange, interacts with a hydrophobic region at the bottom of domain 2 ⁷. This “front-to-back” cis-interaction is also present in our structure (Supplemental Figure 1), providing further circumstantial evidence in favor of its functional importance.

Comparison of type I and type II cadherins

It is informative to compare the process described above for type I cadherins with the adhesion mechanism of type II cadherins revealed in recent structural work ⁸. For all type II cadherin structures, the strand-dimer, including the key Trp2 swapping and N-terminal salt bridge, remains. However, the β strand hydrogen bonds begin from Asn5 in the type II rather than Ile7 in the type I cadherins. A sharp kink at this Asn5 allows the N-terminal polypeptide chain to reach out such that Trp4 joins Trp2 in docking into the acceptor pocket, a characteristic feature of the type II cadherins. As a result, a very large hydrophobic adhesive interface is engaged in type II cadherins, which involves the entire edge of EC1 domain, burying 2700–3300 Å² of accessible surface area ⁸. By comparison, type I cadherins like human E-cadherin use predominantly the first three residues for interaction, and the surface area buried upon dimerization is only about 1600 Å². So far no structure of a type II cadherin in the “closed form” is available. Since, however, there is no Pro5-Pro6 motif in the type II cadherins, strain of the kind seen in the adhesion arm of the type I cadherins is unlikely to exist. It can be envisaged that binding energy gained from the much larger hydrophobic adhesive interface in the type II cadherins may compensate for the activating energy of transforming the intra-molecular docking of Trp2 into inter-molecular docking. It seems that evolution has selected slightly different priming mechanisms for the two types of cadherins: for type I, a Pro5-Pro6 motif that encourages “peeling off” of the adhesion arm and, whereas for type II, the use of a larger hydrophobic interface that facilitates N-terminal swapping.

Concluding remarks

After a type I cadherin precursor undergoes proteolytic activation, the N-terminal segment of EC1 can dock intra-molecularly with the body of EC1 (Figures 3A, 5). The Pro5-Pro6 motif at the base of this segment, however, may introduce strain into this closed conformation that favors dissociation of the N-terminal six-residue “adhesion arm” from the molecular body, priming it for adhesion. When cells juxtapose, forming adherens junctions, the open adhesion arm is stabilized in the adhesive conformation by binding to an opposing cadherin, creating *trans* dimers. Additional inter-molecular interactions are likely to further stabilize the dimer.

The dynamic nature of the adhesion arm may allow a very large group of cadherin molecules to more or less simultaneously form *trans*-interactions. The flexibility of the *trans* dimer may be essential because the extracellular portions of type I cadherin molecules are relatively rigid due to inter-domain calcium-binding, as reported in the five-domain C-cadherin structure ⁷. There is evidence that so-called “*trans*” dimers involving the adhesion arm can also form between cadherins emanating from the same cell surface ^{7;9}. It is possible that the combination of inter-domain rigidity and adhesion arm flexibility provides a geometry that, overall, allows adhesive strand exchange between cadherins on opposing cells. Interactions between cadherins on the same cell might otherwise be overly favored due to their restriction in the two dimensions of the cell surface. Indeed, the configuration of the “*cis*” interactions of desmosomal cadherins observed by electron tomography appeared compatible only with inter-molecular docking of a single adhesion arm, rather than the full mutual strand exchange that could be seen in adhesive dimers ⁹.

Interestingly, type II cadherins may have evolved a distinct means to facilitate dimerization by the strand exchange mechanism. In this case, the formation of a relatively large dimer-specific interface appears to compensate for the entropic penalty of dimerization as well as the activating energy of transforming intra-molecular docking of the adhesion arms into inter-molecular docking. Examination of the sequences of the desmocollins, which retain Trp2 and have a Pro-Ile-Pro motif in place of Ile-Pro-Pro in type I cadherins ³, suggest that they may use a mechanism similar to that of type I cadherins. Desmogleins have Trp2 but a conserved proline motif is absent. It will therefore be of some interest to determine the structures of *trans* dimers of these cadherins, and of representatives of other cadherin subfamilies where residues equivalent to Trp2 apparently do not exist.

Our observations also emphasize the central role that Glu89 or the equivalent residue plays in type I and type II cadherin adhesion, and likely also in desmocollins and desmogleins. This residue forms a salt bridge with the N-terminus of the opposing molecule upon dimerization, where it may compensate for the loss of a similar intra-molecular salt bridge in the monomer. In addition, Glu89 stacks onto the five-membered pyrrole ring of Trp2. Interestingly, Glu89 is also critical for the heterophilic interaction of E-cadherin with the integrin $\alpha_E\beta_7$ ²². In this case, it is unlikely that the presence of the salt bridge is the crucial factor, because E-cadherin monomers can act as ligands for $\alpha_E\beta_7$, and extensions of the N-terminus of E-cadherin do not in themselves prevent $\alpha_E\beta_7$ binding ^{28; 29}. In fact, Glu89 is predicted to lie at the interface between the A domain of α_E and E-cadherin ³⁰. It appears that Glu89 is a critical binding residue for the N-terminus of E-cadherin in both E-cadherin monomers and *trans* dimers, and for $\alpha_E\beta_7$ during heterophilic adhesion. It will be interesting to determine the effect of $\alpha_E\beta_7$ binding on E-cadherin dimerization and signaling, and vice versa, particularly if so-called “*trans*” dimers also form between cadherins emanating from the same cell ^{7;9}.

Thus, the model of the type I cadherin adhesion mechanism presented here provides a useful framework for understanding the homophilic and heterophilic interactions of many cadherin family members. The model also emphasizes the delicate balance between structural elements that is necessary to allow transition from intra-molecular to inter-molecular docking of the adhesion arm, and to favor homophilic cadherin interactions between cells while allowing interactions between cadherins on the same cell surface.

Experimental procedures

Protein production and crystallization

A DNA fragment featuring the first 213 N-terminal amino acid residues (EC1–EC2) of human E-cadherin was obtained by PCR using the full cDNA of the entire human E-cadherin gene as template. The protein-coding DNA sequence contained a point mutation changing Cys9 to Ser9

and was fused at its N-terminus to an initial Met residue, a 6His-tag, a spacer region of 5 amino acids (Ser-Ser-Gly-His-Ile) and the enterokinase recognition site (Asp-Asp-Asp-Asp-Lys). In order to add the entire N-terminal extra sequence to the E-cadherin EC1-EC2 portion, two PCR reactions were performed sequentially. In the first reaction, the 5'-AGCAGCGGCCATATCGACGACGACGACAAGGATTTGGTTATTCCGCCGATC-3' forward primer and the 5'-CCCCCGCTCAGCTCAATCGGTTACCGTGATCAAAATC-3' reverse primer were used. The DNA product thus obtained was then used as template for the second reaction, carried out with 5'-TTTTTCCATGGGCCATCATCATCATCACAGCAGCGGCCATATCGACGAC-3' as forward primer and the same reverse primer used in the first reaction. The final PCR fragment was then cloned into a pET3a vector (Novagen) using the NcoI and BlnI restriction sites.

Overnight protein expression at room temperature in the BL21(DE3)pLysS *E. coli* strain afforded soluble protein in high yield. The cell lysate was then purified on a Ni-column. A second and final purification step using a HiPrep 26/60 Sephacryl S-100 size exclusion column (Pharmacia) was then necessary in order to obtain 100% sample purity as detected by Coomassie staining. The protein sample (buffer TBS pH 7.4, 2mM CaCl₂) was then digested overnight at room temperature with enterokinase (New England Bio-lab) and subsequently passed over a Ni-column to remove all traces of the cleaved 6His-tag as well as all the residual un-cleaved protein. The flow-through of this last affinity chromatography step was collected, dialyzed into 10 mM Tris pH 7.4, 2mM CaCl₂ and brought to a final concentration of 15 mg/ml for crystallization experiments.

Large crystals of the 1-213 fragment of human E-cadherin carrying the Cys9Ser mutation were obtained by the vapor diffusion method at 4° C by mixing a 1 µl drop of the protein sample with an equal volume of 10% PEG 8K, 0.2M CaCl₂, 0.1M Hepes pH 7.5, 5% DMSO solution. Crystals were then frozen in a chemically identical solution to which 20% glycerol was added to serve as cryoprotectant.

Structure determination and refinement

A 2.0Å x-ray diffraction data set was collected at APS using $\lambda = 1.000\text{\AA}$. Diffraction images were processed using the HKL2000 suite³¹ and the structure factors were reduced and scaled with SCALEPACK³¹. Data collection and refinement statistics are shown in Table 1. The structure was determined using the program AMoRe³² and the mouse Ecad-D1D2 crystal structure (PDB code 1EDH) as the search probe. Model building and refinement were carried out using XtalView³³ and Refmac³⁴, respectively. During the final stages, the model was subjected to 5 cycles of TLS refinement³⁴, in which each E-cadherin domain was defined as an independent domain, followed by 10 cycles of maximum likelihood restrained refinement. Water molecules were added manually upon examination of both Fo-Fc and 2Fo-Fc maps. Figures 1 through 4 were generated using PyMOL (<http://www.pymol.org>). The final coordinates have been deposited in PDB data bank with the code: 2O72

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Gary Yue for technical help, and the staff members at APS of Argonne National Laboratories for X-ray data collection. The work was supported by grants from the NIH to JHW and MBB, and from the Crohn's and Colitis Foundation of America to JMGH.

References

1. Gumbiner BM. Regulation of cadherin-mediated adhesion in morphogenesis. *Nat Rev Mol Cell Biol* 2005;6:622–34. [PubMed: 16025097]
2. Cowin P, Rowlands TM, Hatsell SJ. Cadherins and catenins in breast cancer. *Curr Opin Cell Biol* 2005;17:499–508. [PubMed: 16107313]
3. Nollet F, Kools P, van Roy F. Phylogenetic analysis of the cadherin superfamily allows identification of six major subfamilies besides several solitary members. *J Mol Biol* 2000;299:551–72. [PubMed: 10835267]
4. Patel SD, Chen CP, Bahna F, Honig B, Shapiro L. Cadherin-mediated cell-cell adhesion: sticking together as a family. *Curr Opin Struct Biol* 2003;13:690–8. [PubMed: 14675546]
5. Morishita H, Umitsu M, Murata Y, Shibata N, Udaka K, Higuchi Y, Akutsu H, Yamaguchi T, Yagi T, Ikegami T. Structure of the CNR/protocadherin alpha first cadherin domain reveals diversity across cadherin families. *J Biol Chem*. 2006
6. Shapiro L, Fannon AM, Kwong PD, Thompson A, Lehmann MS, Grubel G, Legrand JF, Als-Nielsen J, Colman DR, Hendrickson WA. Structural basis of cell-cell adhesion by cadherins. *Nature* 1995;374:327–37. [PubMed: 7885471][see comments]
7. Boggon TJ, Murray J, Chappuis-Flament S, Wong E, Gumbiner BM, Shapiro L. C-cadherin ectodomain structure and implications for cell adhesion mechanisms. *Science* 2002;296:1308–13. [PubMed: 11964443]
8. Patel SD, Ciatto C, Chen CP, Bahna F, Rajebhosale M, Arkus N, Schieren I, Jessell TM, Honig B, Price SR, Shapiro L. Type II cadherin ectodomain structures: implications for classical cadherin specificity. *Cell* 2006;124:1255–68. [PubMed: 16564015]
9. He W, Cowin P, Stokes DL. Untangling desmosomal knots with electron tomography. *Science* 2003;302:109–13. [PubMed: 14526082]
10. Nagar B, Overduin M, Ikura M, Rini JM. Structural basis of calcium-induced E-cadherin rigidification and dimerization. *Nature* 1996;380:360–4. [PubMed: 8598933]
11. Gallivan JP, Dougherty DA. Cation-pi interactions in structural biology. *Proc Natl Acad Sci U S A* 1999;96:9459–64. [PubMed: 10449714]
12. Ozawa M, Kemler R. Correct proteolytic cleavage is required for the cell adhesive function of uvomorulin. *J Cell Biol* 1990;111:1645–50. [PubMed: 2211831]
13. Koch AW, Farooq A, Shan W, Zeng L, Colman DR, Zhou MM. Structure of the Neural (N-) Cadherin Prodomain Reveals a Cadherin Extracellular Domain-like Fold without Adhesive Characteristics. *Structure* 2004;12:783–805.
14. Haussinger D, Ahrens T, Aberle T, Engel J, Stetefeld J, Grzesiek S. Proteolytic E-cadherin activation followed by solution NMR and X-ray crystallography. *Embo J* 2004;23:1699–708. [PubMed: 15071499]
15. Pertz O, Bozic D, Koch AW, Fauser C, Brancaccio A, Engel J. A new crystal structure, Ca²⁺ dependence and mutational analysis reveal molecular details of E-cadherin homoassociation. *Embo J* 1999;18:1738–47. [PubMed: 10202138]
16. Schubert WD, Urbanke C, Ziehm T, Beier V, Machner MP, Domann E, Wehland J, Chakraborty T, Heinz DW. Structure of internalin, a major invasion protein of *Listeria monocytogenes*, in complex with its human receptor E-cadherin. *Cell* 2002;111:825–36. [PubMed: 12526809]
17. Wang J. Protein recognition by cell surface receptors: physiological receptors versus virus interactions. *Trends Biochem Sci* 2002;27:122–6. [PubMed: 11893508]
18. Chu YS, Eder O, Thomas WA, Simcha I, Pincet F, Ben-Ze'ev A, Perez E, Thiery JP, Dufour S. Prototypical type I E-cadherin and type II cadherin-7 mediate very distinct adhesiveness through their extracellular domains. *J Biol Chem* 2006;281:2901–10. [PubMed: 16253998]
19. Panorchan P, George JP, Wirtz D. Probing intercellular interactions between vascular endothelial cadherin pairs at single-molecule resolution and in living cells. *J Mol Biol* 2006;358:665–74. [PubMed: 16540120]
20. Perret E, Benoliel AM, Nassoy P, Pierres A, Delmas V, Thiery JP, Bongrand P, Feracci H. Fast dissociation kinetics between individual E-cadherin fragments revealed by flow chamber analysis. *Embo J* 2002;21:2537–46. [PubMed: 12032067]

21. Harrison OJ, Corps EM, Kilshaw PJ. Cadherin adhesion depends on a salt bridge at the N-terminus. *J Cell Sci* 2005;118:4123–30. [PubMed: 16118243]
22. Taraszka KS, Higgins JM, Tan K, Mandelbrot DA, Wang JH, Brenner MB. Molecular basis for leukocyte integrin alpha(E)beta(7) adhesion to epithelial (E)-cadherin. *J Exp Med* 2000;191:1555–67. [PubMed: 10790430]
23. Tamura K, Shan WS, Hendrickson WA, Colman DR, Shapiro L. Structure-function analysis of cell adhesion by neural (N-) cadherin. *Neuron* 1998;20:1153–63. [PubMed: 9655503]
24. Morris AL, MacArthur MW, Hutchinson EG, Thornton JM. Stereochemical quality of protein structure coordinates. *Proteins* 1992;12:345–64. [PubMed: 1579569]
25. Bergdoll M, Remy MH, Cagnon C, Masson JM, Dumas P. Proline-dependent oligomerization with arm exchange. *Structure* 1997;5:391–401. [PubMed: 9083108]
26. Rousseau F, Schymkowitz JW, Itzhaki LS. The unfolding story of three-dimensional domain swapping. *Structure* 2003;11:243–51. [PubMed: 12623012]
27. Cailliez F, Lavery R. Dynamics and stability of E-cadherin dimers. *Biophys J* 2006;91:3964–71. [PubMed: 16980367]
28. Corps E, Carter C, Karecla P, Ahrens T, Evans P, Kilshaw P. Recognition of E-cadherin by integrin alpha(E)beta(7): requirement for cadherin dimerization and implications for cadherin and integrin function. *J Biol Chem* 2001;276:30862–70. [PubMed: 11413131]
29. Corps EM, Robertson A, Dauncey MJ, Kilshaw PJ. Role of the alphaI domain in ligand binding by integrin alphaEbeta7. *Eur J Immunol* 2003;33:2599–608. [PubMed: 12938236]
30. Higgins JM, Cernadas M, Tan K, Irie A, Wang J, Takada Y, Brenner MB. The role of alpha and beta chains in ligand recognition by beta 7 integrins. *J Biol Chem* 2000;275:25652–64. [PubMed: 10837471]
31. Otwinowski, Z.; Minor, W. Processing of X-ray diffraction data collected in oscillation mode. In: Carte, CW., Jr; Sweet, RM., editors. *Macromolecular Crystallography*. 276. Academic Press; San Diego, London, Boston, New York, Sydney, Tokyo, Toronto: 1997. p. 307-326.
32. Navaza J. AMoRe: an automated package for molecular replacement. *Acta Cryst* 1994;A50:157–163.
33. McRee DE. XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density. *J Struct Biol* 1999;125:156–65. [PubMed: 10222271]
34. CCP4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr* 1994;D50:760–763.

Figure 1. Ribbon drawing of the strand-dimer structure of human E-cadherin EC1–EC2

The two molecules are related by a crystallographic two-fold axis. Each interacting molecule inserts its Trp2 residue into the hydrophobic acceptor pocket of the dyad-related molecule. Three Ca²⁺ (depicted here as red spheres) are found at the interface between the individual E-cadherin domains, thus imposing rigidity to the molecule.

Figure 2. Detailed interactions between the N-terminal portion of molecule A (magenta) and the hydrophobic acceptor pocket of molecule B (green) within a strand-dimer
 Broken lines show the salt bridge between the N-terminal ^AAsp1-NH₃⁺ and ^BE89 as well as the hydrogen bond between NεH of ^ATrp2 and the carbonyl group of ^BAsp90. The stacking arrangement of residues ^BGlu89, ^ATrp2 and ^BMet92 (see inset on top left corner) is also likely to contribute to the stabilization of this “key-keyhole” interaction observed in all cadherin structures.

Figure 3. Comparison of E-cadherin in closed and adhesive conformations

A, Superposition of the binding pocket of the human E-cadherin (adhesive form) and mouse E-cadherin (closed form; PDB code 1FF5). The human E-cadherin structure reported in this article is shown in green, with its dyad-related molecule depicted in magenta, whereas mouse E-cadherin is shown in yellow. Note the addition of an extra Met to the native N-terminus in the latter structure. The adhesive and the closed forms are well superimposed in this local area. **B**, Superposition of the human E-cadherin EC1–EC2 structure (adhesive form, in green), the human E-cadherin EC1 structure obtained in complex with internalin (closed form, in cyan, PDB code 1O6S) and the mouse E-cadherin EC1–EC2 structure (closed form, in yellow, PDB code 1FF5). The closed structures both have extra residues at the N-terminus (specifically, there is an additional N-terminal Met in the 1FF5 structure and replacement of the native N-terminal Asp1 residue with Gly-Pro-Leu-Gly-Ser1 in the 1O6S structure) so that the N-terminal salt bridge can no longer be formed. In the green and yellow molecules, the P5–P6 motif bulges out prominently, seemingly predisposing the adhesion arm to swing out and promote *trans* strand-dimer formation. By comparison, the cyan molecule's P5–P6 is less protruding, but is in a strained conformation (see text). **C**, Detail of intermolecular interdigitating side chain interactions between the adhesion arm (green) and the body of the opposing molecule (magenta) in the human E-cadherin dimer. **D**, Detail of the adhesion arm of human E-cadherin EC1–EC2 in the adhesive form showing the position of two water molecules (red spheres) bound between backbone atoms of ^AIle4 in the adhesion arm and ^AGlu93 and ^ALeu95 in the body of EC1. Note that only one molecule of the dimer is shown.

Figure 4. Variable conformation of the N-terminal adhesion arm

A, Superposition of the EC1 domains of human E-cadherin EC1–EC2 (adhesive form; green), mouse E-cadherin EC1–EC2 (closed; 1FF5, yellow), *Xenopus* C-cadherin EC1–EC5 (adhesive; 1L3W, light grey) and mouse E-cadherin EC1–EC2 adhesive form (1Q1P, red). The major portion of these molecules is well superimposed. By striking contrast, there is a spectrum of variable swing angles in their N-terminal segments. Such conformational variability suggests that this region confers necessary flexibility to the cadherin molecule to facilitate cell adhesion. **B**, Superposition of the human E-cadherin EC1–EC2 structure (green) and the mouse E-cadherin EC1–EC2 structure (1Q1P; red), highlighting the orientation variation of their strand-dimer counterparts (magenta and yellow, respectively). The calcium ions present in all

structures are omitted for clarity. **C**, Detail of the hydrophobic interactions of Ile4 in the closed form of mouse E-cadherin (1FF5; pale yellow). **D**, As for C but showing the adhesive dimer of human E-cadherin (green and magenta). **E**, As for D but showing the adhesive dimer of mouse E-cadherin (1Q1P; red and yellow).

Figure 5. A model for the binding mechanism of type I cadherins

Starting at the left, the cadherin is synthesized as an inactive pro-form in which docking of the adhesion arm is not favorable due to the absence of a free N-terminus that can form a salt bridge with Glu89 (shown in blue), and perhaps steric hindrance by the prodomain itself. Following irreversible cleavage that removes the pro-sequence, the adhesion arm can dock intra-molecularly into the body of EC1. Intercalation of Trp2 (in green) is stabilized by a stacking interaction with Glu89 and Met92, and the formation of a salt bridge between Glu89 and the free N-terminal amine (shown as dashed red line), among other interactions. This “closed” configuration however, is in equilibrium with an “open” form in which the adhesion arm is dissociated from the body of EC1. The position of this equilibrium may be determined in part by strain in the closed conformation due to the Pro5-Pro6 motif shown in brown. Formation of “adhesive” dimers involves strand exchange with inter-molecular docking of Trp2 and salt bridge formation that is almost identical to that seen in the “closed” form, and the formation of additional inter-molecular interactions also involving the adhesion arm (see text).

Table 1**Data Collection and Refinement Statistics**

Space group	C2
a (Å)	137.34
b (Å)	41.08
c (Å)	59.54
β (°)	111.15
Molecule/asymmetric unit	1
Wavelength (Å)	1.0000
Resolution (Å)	39.1–2.00
Unique reflections	20836
Completeness (%)	98.0 (77.8)
R_{sym}^a (%)	7.3 (32.3)
$I/\sigma(I)$	9.3 (2.5)
Redundancy	6.8 (2.2)
Reflections (work/test)	19768/1068
R^b/R_{free} (%)	19.1/23.4
Ramachandran plot (core/allow/gener/disall)	90.2/9.8/0.0/0.0
Protein/calcium/water atoms	1627/5/245
Average B-factor (Å ²)	19.1
R.m.s.d. from ideal values	
Bond lengths (Å)	0.014
Bond angles (°)	1.445

Numbers in parentheses are for the highest resolution shell

^a $R_{\text{sym}} = \sum_{\text{hkl}} |I - \langle I \rangle| / \sum I$, where I is the observed intensity and $\langle I \rangle$ is the average intensity from observations of symmetry related reflections. A subset of the data (5%) was excluded from the refinement and used to calculate R_{free} .

^b $R = \sum ||F_o| - |F_c|| / \sum |F_o|$