

Structural and energetic determinants of adhesive binding specificity in type I cadherins

Jeremie Vendome^{a,b,c,d,e}, Klara Felsovalyi^{a,b,c,d,e}, Hang Song^{a,b,c,d,e}, Zhongyu Yang^f, Xiangshu Jin^{b,c,e}, Julia Brasch^{b,c,d}, Oliver J. Harrison^{b,c,d,e}, Goran Ahlsen^{b,c,d,e}, Fabiana Bahna^{b,c,d,e}, Anna Kaczynska^{b,c,d}, Phinikoula S. Katsamba^{b,c,d,e}, Darwin Edmond^h, Wayne L. Hubbell^{f,g,1}, Lawrence Shapiro^{b,c,d,1} and Barry Honig^{b,c,d,e,1}

^bDepartment of Biochemistry and Molecular Biophysics, ^cCenter for Computational Biology and Bioinformatics,

^dDepartment of Systems Biology, ^eHoward Hughes Medical Institute, Columbia University, New York, NY 10032

^fJules Stein Eye Institute, ^gDepartment of Chemistry and Biochemistry, University of California, Los Angeles, CA, 90095

^hUniversity of Washington, Seattle, WA 98195

^a J.V., K.F., and H.S. contributed equally to this work.

¹ Correspondence should be addressed to W.L.H. (hubbellw@jsei.ucla.edu), L.S. (lss8@columbia.edu) or B.H. (bh6@columbia.edu)

Abstract

Type I Cadherin cell adhesion proteins are similar in sequence and structure and yet are different enough to mediate highly specific cell-cell recognition phenomena. It has previously been shown that small differences in the homophilic and heterophilic binding affinities of different type I family members can account for the differential cell sorting behavior. Here we use a combination of X-ray crystallography, analytical ultracentrifugation, surface plasmon resonance and double electron-electron resonance (DEER) EPR spectroscopy to identify the molecular determinants of type I cadherin dimerization affinities. Small changes in sequence are found to produce subtle structural and dynamical changes that impact relative affinities, in part via electrostatic and hydrophobic interactions, and in part through entropic effects due to increased conformational heterogeneity in the bound states as revealed by DEER distance mapping in the dimers. These findings highlight the remarkable ability of evolution to exploit a wide range of molecular properties to produce closely related members of the same protein family that have affinity differences finely tuned to mediate their biological roles.

Significance

Type I cadherins comprise a family of cell-cell adhesion proteins that dimerize in a highly specific fashion. There are small differences in dimerization affinities among family members that are evolutionarily conserved and that have profound effects on cell patterning behavior. There are few examples where the molecular origins of small affinity differences between closely related proteins have been explored in depth. We have brought an unusually broad range of technologies to bear on the problem in a unique integrated approach. Our results reveal how a subtle combination of physical interactions combine to tune binding affinities and, in the course of our analysis, we discover a new conformational entropy-based mechanism that can also be exploited by other multi-domain proteins.

Introduction

In metazoans, the elaboration and maintenance of multi-cellular architectures relies upon the ability of cells to specifically adhere to one another. Cadherins constitute a superfamily of single-pass transmembrane proteins that can confer such specific adhesive properties to cells (1). In particular, the classical type I and type II cadherins, which are only found in vertebrates and are characterized by an extracellular region comprised of five extracellular cadherin (EC) domains, have been shown to help drive cell patterning behavior in numerous settings; for example, in morphogenesis (2-4), and in neural patterning (5, 6). Cells expressing the same classical cadherin on their surface generally aggregate through homophilic interactions, while cells expressing different cadherins segregate into distinct layers, which, in at least some instances, remain in contact with each other through heterophilic binding (7-9).

Cell adhesion by classical cadherins is mediated by the dimerization of cadherin extracellular domains emanating from apposed cell surfaces through an interface confined to the N-terminal EC1 domain (Figure 1A). Numerous crystal structures have revealed the atomic details of the *trans* (*i.e.*, between cells) dimerization interface for three type I cadherins: C-, E- and N-cadherins (10-13). In all three cases, the dimer partner molecules swap their N-terminal β -strand (the A*-strand), whose conserved Trp2 residues provide an “anchor” for the adhesive interface by docking into a complementary hydrophobic pocket in the partner protomer (Figure 1A). A second dimerization interface that can form in the *trans* orientation has been observed in crystal structures of mutants of both type I and type II classical cadherins. Specifically, numerous mutations that disrupt strand-swapping in E-cadherin result in the formation of a distinct, lower-affinity homodimer – called the X-dimer due to its appearance – with a binding interface localized around the Ca^{2+} -binding interdomain linker region between EC1 and EC2 (14-16) (Figure 1B). It has been demonstrated that for E-cadherin this interface functions as a kinetic intermediate in the formation of the strand-swapped dimer (16-18) (Figure 1B).

There is a considerable body of evidence demonstrating that the adhesive properties of cells reflect the binding properties of the cadherin molecules they express, and that these properties depend critically on the strand-swapped interface (11, 19-22). Despite their homotypic cell sorting behavior, biophysical studies with purified cadherin ectodomains have shown that cadherins bind both homophilically and heterophilically (9, 23, 24). For the case of E- and N-cadherins, which have slightly different homophilic dimerization free energies: -6.25 kcal/mol for N-cadherin and -5.47 kcal/mol for E-cadherin, the binding free energy associated with heterophilic N-/E- dimerization is intermediate

between these two homophilic values (9). We showed that this combination of homophilic and heterophilic molecular binding affinities predict the observed cell sorting behavior of N- and E-cadherin-expressing cells (9). The link between molecular and cellular behavior depends on the formation of multiple dimers at the contact surface between two adhesive cells, which amplifies the small affinity differences at the level of single molecules (25). It thus appears that subtle differences in the sequence and structure of type I cadherins can have profound effects on cellular behavior, but the molecular origins of these differences have not yet been determined.

Elucidating the source of about 1 kcal/mol difference in binding free energy between two very similar molecules poses a challenging problem. Strand-swapping makes the problem even more complicated since it is not possible to deduce the binding affinity determinants from the crystal structures of the binding interfaces alone. This is because an important consequence of β -strand swapping, or more generally of 3D domain swapping, is that each interaction that stabilizes the dimeric conformation is also formed intra-molecularly in the “closed monomer” conformation (where the N-terminal strand is bound by its own protomer rather than swapping with a partner molecule). As a result, to a first approximation, interactions formed in the dimer must be broken in the monomer so that the net dimerization free energy is due to subtle differences between very similar energetic terms (Chen et al., 2005). We recently found that one source of dimerization energy difference between the swapped dimer and the monomer state of E-cadherin is conformational strain in the A*/A-strand (Figure 1A) in the monomer that is not present in the dimer, thus favoring dimerization (26). This mechanism is likely relevant for N-cadherin as well (27).

Here we report new studies aimed at understanding the relationship between the sequences, structures, and dimerization free energies of type I classical cadherins. We report four new crystal structures of adhesive EC1-EC2 fragments: the P-cadherin swapped dimer, a mutant N-cadherin that reveals its X-dimer structure, and two affinity-mutants of N-cadherin. We use analytical ultracentrifugation (AUC) to quantify the homophilic binding affinities for each type I cadherin, surface plasmon resonance biosensor analysis (SPR) to characterize heterophilic binding between type I cadherin pairs and double electron-electron resonance (DEER) EPR experiments to characterize dimer interactions and dynamics in solution. The combined AUC and SPR measurements provide a nearly complete interaction matrix for this important family of cell adhesion proteins while the x-ray and DEER data make it possible to interpret the affinity measurements in structural and dynamical terms. Our study demonstrates how multiple biophysical and structural approaches can be used in concert to

address mechanistic questions that are not answerable with a more limited repertoire of technologies. Importantly, it clarifies design principles in the type I cadherin subfamily and, in addition, reveals remarkable examples of the fine-tuning of binding specificities for closely related proteins. In this regard, a particularly novel finding is that individual cadherins appear to exist as an equilibrium ensemble of multiple conformational states and that the entropic contribution of this dynamical behavior may have important effects on binding affinities and consequently on cell adhesive specificity.

Results

Homophilic and heterophilic binding affinities of type I cadherins

A phylogenetic tree of type I cadherin EC1 recognition domain sequences reveals two main branches; one containing N- and R-cadherins (henceforth referred to as N-like for convenience of presentation), and a second including E-, C- (E-like) and the more distantly related P-cadherin (28) (Figure 2A). All sequences are from mouse, with the exception of C-cadherin which is unique to *Xenopus*. Nevertheless, C-cadherin clearly groups with E-cadherin. P-cadherin, while closer to E- and C-cadherin in the phylogenetic tree, matches N- and R-cadherin at some residue positions (Figure 2A,B). M-cadherin has a lower sequence similarity with the other type I cadherins and stands by its own on a separate branch.

We produced type I cadherin EC1-EC2 ectodomain fragments of mouse E-, N-, R-, and P-cadherins, and *Xenopus* C-cadherin, using a bacterial expression system. As described in previous work (9), we measured homophilic interactions for each purified cadherin fragment using AUC, and heterophilic interactions using SPR. We have combined these results to produce a relative K_D scale for type I cadherins (Figure 2C, see Figure caption for details). AUC results reveal that N- and R-cadherins each have relatively low homodimerization K_D s, 25.8 and 13.7 μ M respectively, while E- and C-cadherin have significantly higher homophilic K_D s, 96.5 and 126.7 μ M respectively. Though closer in sequence to E- and C-cadherin, P-cadherin has a homophilic binding affinity of 30.9 μ M, similar to that of N- and R-cadherin. AUC measurements for the complete ectodomain of mouse M-cadherin corresponding to EC1-5 revealed a homophilic dimerization K_D of 83.1 ± 4.2 μ M, but the EC1-EC2 fragment was not stable in solution.

SPR data confirmed the AUC results that N- and R-cadherins have strong homophilic interactions (Figure 2C and Supplementary Figure 1). N- and R-cadherins have strong mutual heterophilic binding, and also bind heterophilically with E- and C-cadherins but with lower binding strength. In contrast to N- and R-, the strongest interactions of E- and C-cadherins are heterophilic (with N- and R-cadherins) rather than homophilic (Figures 2C and Supplementary Figure 1). Overall, N- and R-cadherins are found at the high affinity end of the scale for both homophilic and heterophilic interactions, whereas E- and C-cadherins are at the low affinity end of the scale for both types of interactions. Notably, in contrast to its location on the phylogenetic tree, P-cadherin has a homophilic binding affinity in the range of the N-like

cadherins while heterophilically it shows no interaction with N- and R-cadherins but binds weakly to E- and C-cadherins at the concentrations tested.

Structural differences among the strand swapped interfaces of type I cadherins

The crystal structure of an EC1-EC2 ectodomain fragment of mouse P-cadherin at 3.2 Å resolution (Supplementary Figure 2, Supplementary Table 1) forms, as expected, a strand-swapped dimer similar to other type I cadherin adhesive regions. Comparison between type I cadherin EC1-EC2 crystal structures for C-, N-, E-, and now P-cadherins reveals that while the individual EC1 domains superimpose quite well ($C\alpha$ RMSD of EC1 domains < 0.9 Å), in E-like cadherin dimers, and in P-cadherin, the angle between the long axes of the interacting EC1 domains is larger than that in corresponding N-cadherin structures by $\sim 10^\circ$ (85° in E-, 89° in C-, 83° in P-, and 76° in N-cadherin, Figure 3, Supplementary Figure 3, and Supplementary Table 2).

Further analysis of type I cadherin crystal structures offers a potential explanation for this difference. At each swapped interface, residues 2-5, 22, 24, 78 and 92 form a large hydrophobic cluster (Figure 3B-D). Residues 78 and 92, located on the floor of the Trp2 pocket, are subtype specific: residue 78 is Ala in N-like cadherins and Ser in E-like cadherins while residue 92 is Ile in N-like cadherins and Met in E-like cadherins and in P-cadherin. To probe the structural role of positions 78 and 92 we determined the crystal structure of the EC1-2 domain region of the N-cadherin A78S I92M double mutant at 3.2 Å resolution (Supplementary Table 1). (We were unable to crystallize the corresponding E-cadherin S78A, M92I double mutant). The mutant N-cadherin structure reveals a normal strand swapped dimer arrangement. However, as can be seen in Figure 3A and Supplementary Table 2, the EC1-EC1 dimer angle in the double N-cadherin mutant (84°) is much closer to that of E-cadherin than to wild-type N-cadherin. These results suggest that the presence of larger residues at positions 78 and 92 interferes with N-like packing at the swapped interface, leading to an increase of the dimer angle to become E-like.

The effect of positions 78 and 92 on binding affinities is intriguing. The E-cadherin S78A, M92I mutant has a K_D of 23.8 μM as opposed to 96.5 μM in wild type (Table 1) so that changing these two residues transforms E-cadherin into a protein with much stronger N-like homophilic affinity. Surprisingly, the reciprocal mutations in N-cadherin (i.e. N-cadherin A78S, I92M), rather than weakening the binding to mimic E-cadherin, actually increase the binding affinity of N-cadherin from a K_D of 26.8 μM in wild-

type to 4.0 μM in the double mutant (Table 1). Single-site mutants confirm this overall behavior that appears due primarily to position 92 (Supplementary Table 3). Residue 78 has a slight effect on K_D : an alanine in this position makes the binding slightly tighter in each case.

In an attempt to relate the role of positions 78 and 92 on binding affinities we note that the total surface area buried in the interface is approximately 1800 $\text{\AA}^2$ for all crystal structures that have been determined (Supplementary Table 2) but the buried hydrophobic area is significantly larger for N-cadherin ($\sim 1200 \text{\AA}^2$) than for E-, C- and P-cadherin ($\sim 1000 \text{\AA}^2$). Differences in hydrophobic packing are evident in Figure 3B and 3C. However, it is necessary to consider the closed monomer as well as the swapped dimer if one wishes to relate structural features, even qualitatively, to binding affinities in domain-swapped proteins. We have built models of the closed monomers for E-, C-, P-, N- and A78S, I92M N-cadherin (see Methods) and calculated changes in buried area between closed monomers and swapped dimers. As can be seen in Table 2, there is some variability in the increase in buried hydrophobic area upon dimerization for the wild type proteins and indeed the strongest binders (N- and P-) have a greater increase than the weaker ones (E- and C-). However the high affinity N-cadherin A78S, I92M mutant, which has less buried hydrophobic area than N-cadherin itself, can obviously not be accounted for by this effect. Given this and the uncertainty associated with the structures of the closed monomers which depend on model accuracy, we sought other explanations of affinity differences. In the next section we consider the possible role of electrostatic interactions.

The contribution of polar interface residues to dimerization affinities

We analyzed cadherin sequences and structures for subtype-specific charged residues near the dimer interface. Residue 27 is Asn in E-cadherin and Asp in N-cadherin, while residue 90 is Asp in E-cadherin and Asn in N-cadherin (Figure 2B). Switching these residues between N- and E-cadherin shows that the E-like residues (Asn27, Asp90) yield tighter binders than the N-like residues (Asp27, Asn90) in both E- and N-cadherin (Table 1). Remarkably, the D27N, N90D N-cadherin mutant has a K_D that is almost two orders of magnitude smaller than wild type (0.64 μM vs. K_D of 25.8 μM).

In order to determine the relationship between dimer structure and the effect of charged/polar residues at positions 27 and 90, we made N- and E-cadherins with E-like and N-like structure-determining pocket residues at positions 78 and 92, and E-like and N-like polar/ionizable residues at positions 27 and 90 yielding four different combinations. As can be seen in Table 1, E-cadherin with N-like polar residues has a weaker K_D (factor of 2) than wild-type (compare rows 5 and 1) while E-cadherin

with both an N-like conformation and N-like polar residues has an even weaker K_D (factor of 5 compared to wild type; compare row 7 to row 1). Seen in another way, comparing rows 7 and 5 reveals that N-like polar residues weaken the K_D of E-cadherin by a factor of 2 in an E-like conformation (compare rows 6 and 1) but by about a factor of 20 in an N-like conformation (compare rows 7 and 5). A similar trend is observed for N-cadherin. N-cadherin with E-like polar residues has a K_D about 40 times stronger than wild type (compare rows 9 and 1) while N-cadherin with an E-like conformation and E-like polar residues has a K_D about 6 times stronger than wild type (compare rows 10 and 1). Thus, *within* the separate sets of E- and N-cadherin proteins, the highest affinity homodimerization was observed for those containing E-like polar residues (N27 D90) and an N-like pocket (A78 I92), while lowest affinity binding was observed for those containing N-like polar residues (D27 N90) and an N-like pocket (A78 I92) (Table 1). These results strongly suggest that the higher affinity observed for the N-cadherin pocket mutant A78S I92M relative to wild-type is due to more favorable polar interactions in the E-like conformation (compare row 8 to row 1 and row 10 to row 9, in Table 1). It would be interesting to elucidate the structural basis of these observations but reliable calculations would require accurate structural models of both monomer and dimer conformations and long simulations of uncertain accuracy. In the absence of this information, we simply note that the data for both N- and E-cadherin and their mutants can be consistently explained by assuming that E-like polar interactions involving residues 27 and 90 are more favorable than N-like interactions and that the magnitude of these interactions is greater in N-like conformations. We speculate that this latter observation may be due to the more open E-like conformation .

Although our results provide an internally consistent picture of the effects of these four residues, we remain with the paradox that mutants of E- and N-cadherin with virtually identical swapped interfaces consistently exhibit a difference of ~ 2 kcal/mol in binding free energy (compare rows 5, 1, 6, 7 to rows 9, 10, 8, 1 respectively, in Table 1). As a next step in understanding the source of the affinity difference between N- and E-cadherins, we studied the properties of a second *trans* binding interface, the X-dimer (16, 29). Specifically, we asked whether the X-dimer, which acts as a kinetic intermediate that facilitates swapped dimer assembly (16), could also play a role in tuning overall affinities.

X-dimer binding affinities of type I cadherins: The N-cadherin X-dimer is unexpectedly strong

We have previously shown that all strand swap-impaired mutants of E-cadherin form a low affinity X-dimer (Table 1) (16). We produced W2A mutants of the type I N-, R-, C-, and P-cadherins and

determined their homodimerization affinities by AUC (Table 1). Surprisingly, the W2A mutants of N- and R-cadherin have affinities similar to those of the corresponding wild type proteins while homodimerization of C- and P-cadherin mutants was not detectable. X-dimer incompetent K14E or R14E mutants of P-, C-, N-, and R-cadherins all dimerize with essentially wild type affinities (Table 1) as we have previously shown for E-cadherin, indicating that perturbing the X-interface does not affect dimerization affinity mediated by the strand-swapped interface (16, 17). The N-cadherin W2A R14E double mutant, that cannot form either a swapped- or X-dimer, is monomeric in AUC (Table 1) showing that dimerization observed in the W2A mutant is attributable to the X-dimer interface as was observed for E-cadherin (16). We used SPR to assess binding of the X-dimer-incompetent R14E mutant of N-cadherin on a short time scale and found, consistent with results for E-cadherin that no binding was detected between R14E analyte and surface (Supplementary Figure 4A). Thus, as for E-cadherin, the X-dimer of N-cadherin functions as a kinetic intermediate.

Our finding of a high affinity N-cadherin X-dimer contradicts previous reports of a weak X-dimer in N-cadherin suggested by size exclusion chromatography results (27). To clarify this discrepancy, we produced additional N-cadherin mutants designed to disrupt the swapping interface (N-terminal DW-deletion, N-terminal AA-extension, AA-insertion between residues 2 and 3, E89A and W2F). The K_D s of these five mutants are reported in Table 1. Similarly to what we observed for N-cadherin W2A, these mutations either left the N-cadherin homodimerization affinity essentially unchanged (DW-deletion and AA-extension) or even strengthened it (AA-insertion, E89A and W2F) (Table 1). In addition, we assessed homodimerization of N-cadherin W2A mutant by SPR, which gave results consistent with AUC (Supplementary Figure 4B).

We determined a 2.7 Å resolution structure of the N-cadherin AA-insertion mutant, which reveals a novel dimer structure. The overall configuration closely resembles that of the E-cadherin X-dimers, with an rmsd of 1.6 Å when superposed (Figure 4A and 4B). Notably, this mutant swaps A-strands despite also adopting an X-dimer configuration. We had previously pointed out that swapping and X-dimer formation are geometrically incompatible (16). However, the extended A-strand in this mutant allows both X-dimer and strand swap interfaces to engage simultaneously (Figure 4A). In line with its double (swapped plus X-dimer) interface, this mutant dimerizes with a K_D of 3.4 ± 1.7 μM. The presence of an X-dimer in N-cadherin with a comparable binding affinity to that of the strand-swapped dimer explains much of the mutant data in the sense that both interfaces have to be disrupted before a significant reduction in binding affinity can be observed. This suggested the possibility that interplay of

the two interfaces might also help to explain the enhanced binding affinities of the E89A and W2F mutants.

To investigate this possibility, we determined the crystal structure of the N-cadherin W2F mutant at 2.1 Å resolution (Figure 4C, Supplementary Table 2). This protein forms a strand swapped interface whose geometry closely resembles that of the wild-type (Figure 4C and Supplementary Figure 3). However, the W2F R14E double mutant that additionally disrupts the X-dimer interface shows a significantly weaker affinity of $381 \pm 39 \mu\text{M}$ (Table 1). This observation strongly indicates the likelihood that the X-dimer and strand swapped dimer are in equilibrium in the W2F mutant. To probe the possibility of such an equilibrium for W2F and other mutants we turned to EPR spectroscopy.

DEER distance measurements show exchange of swapped and X dimers in N-cadherin mutants

Double electron-electron resonance (DEER) spectroscopy (30) was used to map distributions of inter-spin distances between nitroxide side chains introduced at specific locations on the surface of cadherin EC1-2 *via* site-directed spin labeling (see Methods). This allowed us to infer the dimer conformation and to reveal the flexibility of the EC1-EC2 interdomain linker region of E-cadherin in solution as revealed by structural heterogeneity (Figure 5). We first employed the rigid side chain designated RX (Supplementary Figure 5), which forms cross links between cysteine residues introduced at residues *i* and *i*+2 in a beta strand (31). Figure 5A shows models of 73/75RX in EC1 (label 1) and 114/116RX in EC2 (label 2); Figure 5B shows the corresponding interspin distance distribution determined with DEER. Remarkably, the two well-resolved peaks centered at 40 and 45 Å are in complete agreement with expectations based on X-ray structures, where the 40 Å distance corresponds to label 1 on different protomers when a swapped dimer is formed, and the 45 Å distance to label 1-label 2 located on a single protomer (whether it is monomeric or dimeric) (Figure 5A).

We also tested the effect of removing Ca^{2+} on measured interspin distances. In the absence of rigidification due to Ca^{2+} it is generally assumed that two EC domains in a single protomer rotate freely in solution. Indeed, the interspin distances in the absence of calcium are very broad, spanning from 22 Å to above 60 Å (Figure 5C). Two clusters can be discerned from the distance distribution, one between 32 and 54 Å, and the other centered around 62 Å. The presence of clearly identifiable peaks suggests some degree of order within a large pool of conformations. Nevertheless, this wide range of distances is

indicative of a very flexible interdomain linker region, consistent with our expectations for the calcium-free state.

In order to probe conformational equilibria between Ca^{2+} -bound states in different cadherins, we employed the nitroxide side chain designated R1 (32); unlike RX, R1 only requires a single cysteine for introduction (see Methods and Supplementary Figure 6). In E-cadherin, the native cysteine at position 9 was mutated to serine to ensure single labeling. R1 was introduced at the single position 135 in otherwise wild type and mutant proteins (Figure 6A). In this case, monomers will be invisible in the DEER experiment. However, a dimer will contain two spin labels, and the resultant 135R1-135R1 inter-spin distances predicted based on x-ray structures are sufficiently different in strand-swapped and X dimers to be used as markers for the presence of one conformer or the other: $\sim 58 \text{ \AA}$, $\sim 64 \text{ \AA}$, and $\sim 63 \text{ \AA}$ in the swapped dimer of N-, E-, and P-cadherin, respectively, and $\sim 37 \text{ \AA}$ in both the N- and E-cadherin X-dimers (Figure 6A). Thus, DEER distances can be used to simultaneously monitor the presence of both dimer structures.

Distance probability distributions for wild-type N-, E-, and P-cadherin as well as mutant N- and E-cadherin EC1-2 domains are shown in Figure 6B. Distance determinations by DEER in the range beyond about $70 \text{ \AA}$ are challenging. Such long distances are encountered in the distributions shown in Figure 6, where the dotted traces identify parts of the distribution beyond the limits for quantitative determination set by the data collection time, which in the case of the cadherins is about $70 \text{ \AA}$ (see Methods). Nevertheless, the high quality data reveals the existence of populations beyond these limits, but the positions and widths are undetermined. The N-cadherin wild-type protein shows a broad inter-spin distance distribution with a shoulder at $55 \text{ \AA}$ and a peak located at $63 \text{ \AA}$. The data reveal populations beyond the $70 \text{ \AA}$ limit that are indicated in Fig. 6 with dashed traces. The shoulder and the first peak are close to the $58 \text{ \AA}$ average distance predicted from the crystal structure of N-cadherin. (Figure 6B). Note that this DEER distance distribution of wild type N-cadherin is very similar to that of the X-dimer-incompetent R14E mutant, which forms only a strand swapped dimer (Figure 6B). These results suggest that both the wild-type and the R14E mutant of N-cadherin primarily adopt a swapped dimer conformation but that there is a considerable structural heterogeneity associated with this conformation. In contrast, wild-type E-cadherin shows a major peak at $70 \text{ \AA}$ (Figure 6B), consistent with its larger dimer angle than the average N-cadherin swapped dimer (Figure 3A). The width of the overall distribution is also narrower than that in wild-type N-cadherin, suggesting that N-cadherin has a broader

distribution of dimer angles in solution than E-cadherin. Note that P-cadherin has an N-like distance distribution (Figure 6B). We will return to the implications of this observation below.

Importantly, the lack of inter-spin distances around 37 Å indicates the absence of X-dimers in all wild type N-, E-, and P-cadherins, consistent with x-ray structures. In contrast, the N-cadherin W2A mutant, as well as the AA insertion mutant show a single narrow peak centered at 38 Å (Figure 6B), consistent with the predicted distance for the X-dimer conformation. For the N-cadherin W2F and E89A mutants, which dimerize with higher affinity than wild type (Table 1), the DEER distance distributions show both an X-dimer peak at 38 Å, and three distinct peaks between approximately 44 and 70 Å which we associate with swapped dimers (Figure 6B). Thus, consistent with indications from AUC data (see above) in these two mutants the X- and swapped dimer conformations appear to be in equilibrium with one another.

The shape of the distance distributions is also informative. The X-dimer peak in both the W2F and E89A mutants is narrow while the swapped-dimer peaks are distributed over a wide range from $\approx 44 - 70$ Å. In the N-cadherin tryptophan pocket mutant A78S I92M, the long distance peaks starting at 43 Å appear intermediate between the wild-type and the other two tight binders W2F and E89A, while no peak corresponding to the X-dimer distance (~ 38 Å) is evident (Figure 6B). This distance distribution is quite similar to that of the E-cadherin pocket mutant S78A M92I, which has three peaks that span a greater distance range than that for wild-type E-cadherin (Figure 6B). This indicates that the pocket mutations not only affect the structure of E-cadherin but also its dynamical properties.

Discussion

Cell surface adhesion molecules such as cadherins generally appear as members of families of closely related proteins that carry out similar functions, most notably cell-cell recognition. Type I cadherins bind to one another both homophilically and heterophilically, with a range of affinities that can vary over about two orders of magnitude. These affinities are evolutionarily conserved among species (9) and arise from small changes in sequence that might be expected to have little effect on affinities since the relevant amino acid substitutions are often conservative (Figure 2B). Our primary goal in this study has been to elucidate the sequence, structural and energetic origins of the small differences in cadherin dimerization affinities that underlie their biological specificity.

We have determined that the N-like, N- and R-cadherins have both greater homophilic and heterophilic binding affinities than the E-like E- and C-cadherins, while P-cadherin, which is more E-like

in sequence, exhibits atypical behavior (Figure 2C). The clearest structural difference between N-like and E-like cadherins is in the EC1-EC1 dimer angle that is larger in the latter than in the former. The more open E-like conformation is seen in the crystal structures of C- and P-cadherin (which both contain a Met at position 92) and in the A78S, I92M mutant of N-cadherin where the Trp pocket residues corresponds to those of E-cadherin. The larger dimer angle in E-like cadherins is also evident from the main peak in DEER distance distributions of wild type N-, P- and E-cadherin. The DEER distance distributions of N- and P- cadherin span a greater distance range than that of E-cadherin suggesting that the swapped dimer of N- samples a wider range of dimer angles which in turn points to greater protein motion on the domain level (see discussion below).

A second difference between N-like and E-like cadherins involves their X-dimers. E-cadherin mutants such as W2A that ablate strand-swapping have dimerization affinities substantially weaker than wild type. In contrast, the DEER distance distribution of the W2A mutant of N- cadherin shows that it forms an X-dimer, and AUC shows that it has an affinity similar to that of wild type (Table 1). Thus, one would expect both the swapped and the X-dimer species in equilibrium as seen in the DEER distance distributions of W2F and E89A but, in fact, only the swapped dimer is observed for the wild-type N-cadherin. This implies, that for a still unknown reason, the X-dimer affinity of wild type N-cadherin is actually weaker than that of W2A and other mutants which form stable X-dimers (K_D s in the range of 25-50 μ M, Table 1). Crucially however, as is the case for E-cadherin (as well as the type II cadherin-6) the X-dimer in N-cadherin functions as a kinetic intermediate on the path to the formation of the mature strand-swapped interface (16, 17). This can be deduced from the observation that the X-dimer-incompetent R14E mutant doesn't form strand-swapped dimers in SPR experiments where there is limited time to interact (Supplementary Figure 4A) but does in AUC experiments. This behavior is entirely consistent with that observed for E-cadherin (16,17).

Our attempt to probe the structural and dynamic tuning of binding affinities with site-directed mutants answered numerous questions, but also raised others. Indeed, every experiment carried out with the goal of designing an E-cadherin mutant with N-like properties was successful. In contrast, our attempts to design N-cadherin mutants with E-like affinities failed. Replacing the Trp pocket residues with those of E-cadherin (N-cadherin A78S, I92M) which was intended to weaken dimerization, produced an N-cadherin "super mutant" with a low K_D of ~ 3.0 μ M. Similarly, replacing polar residues 27 and 90 to those of E-cadherin (N-cadherin D27N, N90D) instead of weakening dimerization produced an even stronger super mutant ($K_D \sim 0.6$ μ M). Other mutations known to weaken binding in E-cadherin

(W2F, E89A and the AA insertion mutant) all yielded N-cadherin super mutants. How can these apparently conflicting observations be resolved?

The high affinity of the AA insertion double mutant appears straightforward to explain as it is consistent with its forming a large combined interface and indeed, its buried surface area ($\sim 2000 \text{ \AA}^2$) is greater than any other type I cadherin. However, the high affinity of the N-cadherin W2F and E89A mutants cannot rely on the simultaneous formation of both interfaces in a single structure since both the X-dimer peak and longer distance swapped-dimer peaks are observed in the DEER distance distribution of the W2F and E89A mutants (Figures 6B), indicating two distinct types of conformations. This in turn suggests an entropic contribution to the dimerization affinities of both mutants and is consistent with an equilibrium involving both the X-dimer and swapped dimer.

Summarizing the analysis of the mutant data, we are able to account qualitatively for the affinity of every N- or E-cad mutant relative to wild type based on structural changes, polar interactions, and the presence of multiple conformations. We are now in a position to analyze the differences among the wild type proteins themselves, particularly with respect to the greater dimerization affinities of N-like and P-cadherin relative to E-like cadherins. As discussed above, differences in buried hydrophobic area may play a significant role (Table 2). In addition, we note that the stronger binding N- and P-cadherins show wider distance distributions than the weaker E-cadherin, and that mutations that confer greater affinity are associated with even wider distance distributions (Figure 6B). The DEER data thus suggest that greater conformational freedom and the presence of multiple dimeric conformations result in more favorable entropic contributions to binding.

In principle, numerical values for configurational entropy differences could be determined from normalized distance probability distributions by taking each discrete distance measured in the DEER experiment as a microstate i with a corresponding probability p_i and computing S according to $-R\sum_i p_i \ln(p_i)$. However, for the long-range distances measured here the uncertainty in the widths and extent of the distribution beyond $70\text{\AA}$ precludes a reliable numerical calculation. Nevertheless, it is qualitatively clear that both the N- and P-cadherin dimers have greater configurational entropy relative to E-cadherin. Entropic stabilization due to an equilibrium between multiple conformations has not, to our knowledge, been posited previously, although similar effects have been discussed in the context of the binding of natively unstructured peptides (33).

We recognize that the explanation of binding affinities presented here is primarily qualitative in nature. Indeed, the complexities of the systems involved would more than challenge the current state of detailed binding free energy calculations. However, the large body of data and the accompanying analysis offer a far more complete picture of the physical principles underlying cadherin function than has been available previously, and they reveal new ways in which evolution can exploit fundamental physical principles in the subtle tuning of binding affinities. In particular, the entropic contribution resulting from differences in the extent of dimer flexibility constitutes a general mechanism that is available to multi-domain proteins and which we have shown previously to also play a role in the binding of cadherins located in the two-dimensional environment of cell surfaces (34).

Perhaps most notably, our study illustrates how evolutionary fine-tuning can proceed in ways that defy the simple logic of experiments that probe local regions of protein structure. An intriguing example is provided by the identity of the residues at positions 27 and 90 that, when corresponding to those of wild type N-cadherin, actually lower affinities relative to those of E-cadherin. Overall, N-cadherin has been designed to have a greater dimerization affinity than E-cadherin, but these locations have apparently been used to ensure that this binding is not too tight. More generally, type I cadherin dimerization affinities seem to be coded on the entire EC1 structure and via its dynamic properties. This contrasts to other families of adhesion proteins, such as nectins, for which intra-family specificity can be mapped to individual residues (35).

Accession Codes

Protein Data Bank: coordinates for mouse P-cadherin EC1-EC2 wild-type, mouse N-cadherin EC1-EC2 A78S I92M, AA insertion, and W2F have been deposited with accession codes 4NQQ, 4NUM, 4NUP and 4NUQ, respectively.

Acknowledgements

This work was supported in part by the US National Institutes of Health (R01 GM062270 to L.S., R01 EY005216 and P30 EY00331 to W.L.H.), the National Science Foundation (MCB-0918535 to B.H.), and the Jules Stein Professorship Endowment (W.L.H.). Use of the National Synchrotron Light Source, Brookhaven National laboratory (BNL), at the X4A and X4C beamlines was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-98CH10886; the beamlines are operated by the New York Structural Biology center. We thank J.

Schwanof and R. Abramowitz at BNL for support with synchrotron data collection, and C. Altenbach for his help with EPR distance data analysis. K.F. was supported by a training Grant T32 GM082797 from the National Institute of Health. D.E. was supported by SPURS.

Author Contributions

Conception and design: J.V., K.F., H.S., W.L.H., L.S., B.H.

Acquisition of data: J.V., K.F., H.S., Z.Y., X.J., J.B., O.J.H., G.A., F.B., A.K., P.S.K., D.E.

Analysis and interpretation of data: J.V., K.F., H.S., Z.Y., X.J., J.B., O.J.H., G.A., P.S.K., W.L.H., L.S., B.H.

Drafting or revising the article: J.V., K.F., H.S., J.B., O.J.H., W.L.H., L.S., B.H.

Competing Financial Interests

The authors declare no competing financial interests.

Materials and Methods

Phylogenetic tree

Sequences of mouse E-, P-, N-, R- and M-cadherin and *Xenopus laevis* C-cadherin EC-1 domains (residues 1-101) were aligned using Muscle (36), and the sequence alignment obtained was used to build the phylogenetic tree using the maximum likelihood method PhyML (37).

Bacterial protein production

Coding sequences of mouse N-, E-, P-, R-cadherin EC1-2 (Asp1-Val216, Asp1-Asp213, Glu1-Asp213, and Asp1-Asp215, respectively) and *Xenopus* C-cadherin EC1-2 (Asp1-Asp213) were amplified by PCR from cDNA libraries (Clontech, Mountain View, CA) and cloned in frame with an N-terminal hexahistidine-tagged SUMO protein into the BamHI/NotI sites of the pSMT3 vector (a modified pET-28b(+)) expression plasmid). Cleavage of SUMO-fusion proteins with Ulp1 (Ubiquitin-like protease 1) after a Gly-Gly motif yields cadherin proteins with native termini. Any extra amino acids occurring after the cleavage site due to cloning were removed by using the QuikChange site directed mutagenesis kit (Stratagene, Agilent Technologies, Santa Clara, CA) to ensure native N-termini of all proteins used in our studies, unless specifically altered. We introduced all point mutations using either the QuikChange mutagenesis kit or KOD Hot Start DNA Polymerase kit (Novagen, EMD Chemicals, Gibbstown, NJ). For SPR analysis, a Gly-Gly-Gly-Cys motif was introduced by PCR before the stop codon in wild-type P-, C-, N-cadherin, E-cadherin W2A, N-cadherin W2A, and N-cadherin R14E mutants, yielding a free C-terminal cysteine in these constructs for attachment to thiol-reactive SPR chips. Biotinylated wild-type E- and N-cadherin proteins were produced with a C-terminal Avi tag as described previously (9) for immobilization on neutravidine-derivatized SPR chips. R-cadherin was fused with BCCP (biotin carboxyl carrier protein) after Asp215 for *in vivo* biotinylation and subsequent chip immobilization. All PCR primers were obtained from Life Technologies (Carlsbad, CA).

For protein production, *E. coli* Rosetta 2DE3 pLysS (Novagen) or BL21-Gold (DE) cells (Agilent) were transformed with prepared pSMT3 vector and grown at 37°C until OD600 reached 0.6 shaking at 200rpm. To induce protein expression, we added 100µM IPTG and lowered the temperature to 18°C. After 18h cells were harvested by centrifugation at 4,000g for 15min. Pelleted bacteria were

resuspended in lysis buffer (500mM NaCl, 10-20mM Tris-HCl pH8.0, 20mM Imidazole pH8.0, 3mM CaCl₂) and lysed for 3-6 minutes by sonication in 15-second intervals with 45 seconds rest in between pulsing. Cell debris was spun down at 4°C and 20,000g for 30min-1 hour and His-tagged proteins were extracted from the clear lysate by flowing over nickel charged IMAC Sepharose 6 Fast Flow resin (GE Healthcare, Pittsburgh, PA). Beads were subsequently washed with 20-40 column volumes of lysis buffer to remove contaminants and 6xHis-SUMO-fusion proteins were eluted with lysis buffer containing 250mM imidazole. The 6His-SUMO tag of the fusion proteins were cut enzymatically by adding Ulp-1 to a final concentration of 2µg/mL to the elution. Proteins were dialyzed into a low ionic strength buffer (100mM NaCl, 10-20mM Tris-Cl pH8.0 and 3mM CaCl₂). We removed Ulp1 (which has a His tag), cleaved 6His-SUMO tags, and any remaining uncut fusion protein by batch binding to nickel charged IMAC resin equilibrated in dialysis buffer. The cadherins were further purified by anion exchange chromatography (Mono Q 10/10 HR, GE Healthcare) using a NaCl gradient, and size exclusion chromatography (HiLoad 26/60 SuperdexTM S75 prep grade, GE Healthcare) in a final buffer of 150mM NaCl, 10mM Tris-Cl pH8.0 and 3mM CaCl₂. Proteins were concentrated to a final concentration of approximately 1-10mg/mL using Amicon Spin concentrators (Millipore, Billerica, MA) and flash frozen. Cysteine-containing proteins were purified using the same method – except all buffers were supplemented with 1mM TCEP to keep cysteines reduced.

Mammalian protein production

Mouse M-cadherin EC1-5 (Ala1-Ala547) was expressed in HEK293 cell lines as secreted proteins with a PTPα signal sequence and C-terminal hexa-histidine tag. Conditioned media were harvested two days after transient transfection using polyethylenimine buffered to pH 7.0 with 1M NaOH, and brought to high salt concentration (500mM NaCl) before purification. The protein was purified using IMAC Sepharose 6 Fast Flow (Ion Metal Affinity Chromatography) resin charged with Nickel and dialyzed against 150mM NaCl, 10mM Tris-HCl pH8.0 and 3mM CaCl₂, followed by two chromatographic steps: ion exchange (MonoQ 10/100GL) and size exclusion (S200 26/60) in a final buffer of 150mM NaCl, 10mM Tris pH8.0 and 3mM CaCl₂.

Crystallization, Data Collection, and Refinement

We expressed and purified mouse P-cadherin EC1-2 as described above and used it for crystallization studies at 2.58mg/mL. Using the vapor diffusion method, protein crystals grew at 20°C after combining 0.6μL protein with 0.6μL well solution composed of 38% (w/v) PEG 6000, 350mM calcium chloride and 100mM Bis-Tris pH6.5. The crystals grew within 48h and were flash frozen in liquid nitrogen after being immersed briefly in cryo protectant (15% (w/v) butane-2*R*,3*R*-di-ol, 18% (w/v) PEG 6000, 350mM calcium chloride, 100mM Bis-Tris pH6.5).

For mouse P-cadherin EC1-2, data was collected on a single frozen crystal at the X4C beam line of the National Synchrotron Light Source, Brookhaven National Laboratory at a wavelength of 0.979Å. We processed the data using the HKL suite (38) and solved the structure by molecular replacement with mouse E-cadherin (PDB code: 2QVF) as search model using Phaser (39). Refinement was carried out by manual building in Coot (40) followed by automated refinement in Phenix (41). Ramachandran plot statistics for the final model are 96.4% favored, 3.6 % allowed and 0% outliers.

Crystals of N-cadherin EC1-2 A78S192M mutant were obtained in 15% PEG3350, 0.2M sodium chloride, 0.1M Tris, pH 8.5. The crystals were cryoprotected with the mother liquor supplemented with 30% glycerol before flash cooling in liquid nitrogen. These crystals belong to space group $P2_1$ with cell dimensions $a = 59.8 \text{ Å}$, $b = 221.2 \text{ Å}$, $c = 72.2 \text{ Å}$, and $\beta = 103.9^\circ$. Crystals of N-cadherin EC1-2 mutant with AA insertion between residues 2 and 3 were obtained in 0.3M magnesium formate, 0.1M HEPES, pH 7.5, and were cryoprotected with the mother liquor supplemented with 30% glycerol for freezing in liquid nitrogen. These crystals belong to space group $P2_12_12$ with $a = 175.5 \text{ Å}$, $b = 65.8 \text{ Å}$, and $c = 102.5 \text{ Å}$. Crystals of N-cadherin EC1-2 W2F mutant were obtained in 4% PEG4000, 0.2M magnesium chloride, 0.1M MES, pH 6.5. The crystals were cryoprotected with the mother liquor supplemented with 30% ethylene glycol before flash cooling in liquid nitrogen. These crystals belong to space group $C2$ with cell dimensions $a = 116.6 \text{ Å}$, $b = 86.2 \text{ Å}$, $c = 46.7 \text{ Å}$, and $\beta = 98.5^\circ$.

Diffraction data for all mouse N-cadherin EC1-2 mutants were collected on single crystals at 100K at X4C beam line in the National Synchrotron Light Source at Brookhaven National Laboratory, and processed with the HKL program suite (38). The structures were solved by molecular replacement with PHASER using the structure of wild type N-cadherin EC1-2 (PDB ID code: 2QVI). Manual rebuilding was done with COOT, and refinement was performed using REFMAC (42) implemented in the CCP4 program suite

(Collaborative Computational Project Number 4, 1994). The statistics of data collection and refinement are summarized in Supplementary Table 1.

Structural biology applications were provided by SBGrid (43).

Calculation of buried surface area

Buried surface areas were calculated using SURFV, which is available from download on the Honig Lab website http://wiki.c2b2.columbia.edu/honiglab_public/index.php/Software (44).

Calculation of EC1-EC1 dimerization angles in swapped dimers

Each EC1 domain was treated as a rigid body consisting of point masses at C α positions of residues 1-100. Other backbone atoms, side chains, and all atoms from the inter-domain linker region (residues 101-104) were excluded from analysis. The three principal axes of inertia of the rigid body were calculated (45). The long axis of the domain was then defined as the principal axis representing the axis of rotation that requires the least amount of torque to stop the rotation of the rigid body. For a cadherin EC domain, whose shape roughly resembles a cylinder, this long axis is analogous to the axis of symmetry along the cylinder. The angle between the two EC1 domains in a swapped dimer was subsequently derived from the dot product of their respective long axes.

Building models of the closed monomers

A structural model of the E-cadherin closed monomer was derived from the crystal structure of the closed monomer of mouse E-cadherin EC1-2 (PDB 1FF5) (15) which contains an extra N-terminal methionine that prevents the formation of a crucial salt-bridge between the N-terminus and Glu89 (19). In the model, the N-terminal methionine was removed and a local minimization carried out with a harmonic constraint with a minimum distance of 5.0 Å between the N atom of the NH3 terminus and the CG atom of Glu89. The local minimization consisted of 100 steps of steepest descent and 300 steps of a conjugate gradient minimization. We verified that a proper salt-bridge between the N atom of the NH3 terminus and the CG atom of Glu89 was formed in our final model. Charmm was used for the constrained minimization (46).

A model for the closed monomer of N-cadherin was built from chain B of the wild-type swapped dimer in the crystal structure of two domains N-cadherin (pdb file 2QVF). All EC1-2 domain residues except for the A*/A-strand (residues 1-10) were assigned their crystallographic coordinates. For the A*/A-strand, a homology model of N-cadherin closed monomer was built for the whole EC1 domain with the program Nest (47) using the E-cadherin closed monomer model described above as a template, for the A*/A-strand residues. The chimeric structure was then locally refined using a constrained minimization with Charmm. Specifically, the minimization consisted of 200 steps of steepest descent algorithm followed by 300 steps of a conjugate gradient minimization. All C α s for residues 11-210 were restrained to their position with a force of 20 kcal.mol⁻¹.Å⁻².

Models of the closed monomer of N-cadherin A78S I92M double mutant and P-cadherin wild type were constructed using the same procedure as for the model of N-cadherin wild type closed monomer: core residues (from residues 11 to the end of the EC2 domain) were assigned their crystallographic coordinates (from PDB 4NUM chain A and PDB 4NQQ chain A for N-cadherin A78S I92M and P-cadherin, respectively)

Analytical ultracentrifugation

Sedimentation equilibrium measurements: Analytical ultracentrifugation (AUC) equilibrium experiments were performed at 25°C, using a Beckman XL-A/I ultracentrifuge equipped with a Ti60An rotor. Data was collected using UV absorbance at 280 nm. Samples were dialyzed in TRIS 10 mM, NaCl 150 mM, pH 8.0 for 16 hours at 4°C and loaded into six-channel equilibrium cells with parallel sides and sapphire windows. For Cys-containing proteins 1 mM TCEP was added to buffer. 120mL aliquots of sample diluted to 0.7 (30), 0.46 (20) and 0.24 (10) mg/mL (mM) were loaded, respectively, into three channels A, B and C of the cell, with three of the channels used for buffer reference. Samples were spun at 15000 rpm (16350*g) for 20 hours, after which four scans were collected at a rate of 1 per hour. The rotor speed was then increased to 19000 rpm (26230*g) for 10 hours, after which four additional scans were collected at the same rate. The speed was further increased to 23000 rpm (38440*g) for another 10 hours and four more scans were recorded under the same conditions. During the last step, the rotor speed was increased to 27000 rpm (52970*g) for four more scans, resulting in a total of 16 scans for each concentration and a total of 48 scans per protein. Each experiment was reproduced at least twice.

The data was processed and analyzed using HeteroAnalysis 1.1.44 software (<http://www.biotech.uconn.edu/auf>) and buffer density and protein v-bars were calculated using the SednTerp software (Alliance Protein Laboratories, San Diego, CA). The data for all concentrations and speeds were globally fit using nonlinear regression to either a monomer-homodimer equilibrium model or ideal monomer model.

SPR binding assays

Binding assays were performed using a Biacore T100 biosensor equipped with a Series S CM4 chip (GE Healthcare). Biotinylated versions of mouse E-, N- and R-cadherins EC12 domains were captured over neutravidin-immobilized surfaces. Mouse P- and N-cadherins EC12, together with *Xenopus* C-cadherin EC12 were covalently immobilized via a C-terminal cysteine using a ligand thiol-coupling protocol. The biotinylated and thiol-coupled N-cadherin surfaces yielded similar results suggesting that the tethering method does not influence the binding responses. Each cadherin ligand was tethered to the chip surface at 70.0 mM monomer concentration, which was calculated using the homophilic binding K_D s as listed in table 1.

Neutravidin immobilization and capture of biotinylated cadherin were performed as described previously (9). Thiol-coupled proteins were immobilized in HBS pH 7.4 (10 mM HEPES, 150 mM NaCl, pH 7.4), 3 mM CaCl_2 at 25°C using a flow rate of 20 $\mu\text{L}/\text{min}$. For the immobilization reaction, the carboxyl groups were activated for 2 minutes using 400 mM EDC (*N*-ethyl-*N*-(3-dimethylaminopropyl)carbodiimide), mixed at 1:1 ratio (v/v) with 100 mM NHS (*N*-hydroxysuccinimide). Subsequently, a solution of 120 mM PDEA, was mixed with 0.1 M sodium borate pH 8.5 at 2:1 ratio (v/v), to yield a final concentration of 80 mM PDEA and injected over the same flow cell for 4 minutes. A sample of the cadherin protein to be immobilized was freshly desalted in 10 mM sodium acetate, pH 4.0 and sequentially injected over the activated surface at 10-50 mg/mL, depending on the protein, until the desired immobilization level was achieved. Any remaining disulfides were blocked using a four-minute injection of 50 mM L-cysteine/1.0 M NaCl in 0.1M sodium acetate, pH 4.0.

Cadherin binding experiments were performed at 25°C in a running buffer of 10 mM TRIS-HCl, pH 8.0, 150 mM NaCl, 3mM CaCl_2 , 0.25 mg/mL BSA and 0.005% (v/v) Tween 20. Soluble cadherin (analytes) were diluted in running buffer to a 12.0 mM monomer concentration, which were calculated using the homophilic K_D values listed in table 1. Samples were injected for 60s at 50 mL/min followed by a 60-s

dissociation phase. A one-minute buffer wash at the end of the binding cycle minimized sample contamination of the fluidics system of the instrument. Each cadherin was tested at least twice to verify the reproducibility of the assay and confirm stability of the immobilized surfaces. The binding responses were double-referenced (9) and processed using Scrubber 2.0 (BioLogic Software, Campbell, Australia).

EPR analysis

Spin labeling of cysteine mutants: All labeling reactions were carried out in 25mM MOPS pH 6.8, 100mM NaCl, and 3mM CaCl₂. In a typically labeling reaction, a 5- to 10-fold molar excess of S-(2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate (MTSL; Toronto Research Chemicals, Toronto, Canada) was added to the protein (5-25 μ M) immediately after buffer exchange. For labeling with RX, a 1:1.1 molar ratio of double cysteine to HO-1944 (Toronto Research Chemicals) was used. The reaction was allowed to proceed at room temperature for 30 min-1 h and then at 4°C overnight. Excess label was removed by washing and the protein concentrated in an Amicon concentrator (Millipore) 10mM Tris-HCl pH 8.0, 150mM NaCl, and 3mM CaCl₂. The extent of labeling was assayed with Aldrithiol™-4 (Sigma-Aldrich, St. Louis, MO) following established protocols (48). For RX-labeled proteins, the final buffer conditions are 25 mM HEPES pH 7.4, 100 mM KCl, 3 mM CaCl₂, and 10% glycerol. For calcium-free experiments, a small amount of RX-labeled E-cadherin was dialysed against 25 mM HEPES pH 7.4, 100 mM KCl, 10% glycerol, and 5 mM EGTA overnight. Except for the N-cadherin W2A mutant, R1-labeled protein was washed and concentrated in buffer made with D₂O. The final D₂O content of the solvent water was estimated to be 95%. The predicted average R1-R1 interspin distances as well as the distance distributions based on favorable rotamer conformations were calculated using the PRONOX program's webserver (<http://rockscluster.hsc.usc.edu/research/software/pronox/pronox.html>) (49) for proteins with available crystal structures. The RX labels were modeled onto the crystal structure of bidomain E-cadherin (PDB code 2QVF) using Discovery Studio (Accelrys Software, San Diego, CA), and the interspin distances measured in PyMol (Schrödinger, Portland, OR). Positions for spin label attachment are chosen such that they are not expected to interfere with functional interfaces or tested mutation.

DEER Spectroscopy: Experimental procedures for the four-pulse DEER experiments on the spin-labeled cadherin mutants were published previously (50). Protein concentrations for DEER experiments varied between 100 and 400 μ M depending on the solubility of the mutants. For each measurement 12- 16 μ L

sample with 25 % v/v glycerol as cryoprotectant were loaded into a quartz capillary tube (1.5 ID× 1.8 OD; VitroCom, Mountain Lakes, NJ) and then flash-frozen in liquid nitrogen. All DEER measurements were performed at 80 K on the Bruker ELEXSYS 580 equipped with a Q-band resonator (ER5107DQ), a SuperQFTu-EPR bridge and a 10 W Q-band amplifier. The $(\pi/2)$ and π pulses were adjusted to be exactly 16 and 32 ns, respectively. A two-step phase cycling (+x, -x) is carried out on the first $(\pi/2)$ pulse. Time domain signal collected for each sample varied from 3 to 5.5 μ s depending on the expected distance(s). The longer time domain of 5.5 μ s was achieved using deuterated buffer and d_8 -glycerol (99%, Cambridge Isotope Laboratories, Tewksbury, MA). Signal acquisition time varied from 6 to 16 hours depending on sample concentration. Data were analyzed using the program LongDistances (51). The upper limit of accurate mean distance and width determination are calculated using the equations $r_{max,<r>} \approx 5\sqrt[3]{t_{max}/(2\mu s)}$ and $r_{max,\sigma} \approx 4\sqrt[3]{t_{max}/(2\mu s)}$ (30). For our acquisition time of 5.5 μ s, $r_{max,<r>} \approx 70$ Å and $r_{max,\sigma} \approx 56$ Å.

Tables

Table 1: Dissociation constants (K_D) for wild type and mutant type I cadherins^{a,b}. Dissociation constants from previous studies are indicated. Data are given as mean $\pm$ s.d.

	Mutation	K_D (μ M)					
		E-cadherin	N-cadherin	C-cadherin	R-cadherin	P-cadherin	M-cadherin
1	Wild-type	96.5 $\pm$ 10.6 ¹	25.8 $\pm$ 1.5 ¹	126.7 $\pm$ 19.7	13.7 $\pm$ 0.2	30.9 $\pm$ 1.0	83.1 $\pm$ 4.2
2	W2A	916 $\pm$ 47 ²	26.9 $\pm$ 3.1	monomer (>1000)	11.8 $\pm$ 0.59	monomer (>1000)	n.d.
3	K/R 14E	117 $\pm$ 8 ²	41.4 $\pm$ 2.11	122 $\pm$ 2.5	10.3 $\pm$ 2.25	38.3 $\pm$ 2.18	n.d.
4	W2A, K/R14E	monomer ²	monomer ²	n.d.	n.d.	n.d.	n.d.
	Mutation	K_D (μ M)		Mutant description			
5	S78A, M92I	23.8 $\pm$ 1.6	-	E-cadherin with N-like Trp2 pocket lining residues			
6	N27D, D90N	191 $\pm$ 24	-	E-cadherin with N-like polar residues around Trp2 pocket			
7	S78A, M92I, N27D, D90N	506 ^c	-	E-cadherin with both N-like polar residues and N-like Trp2 pocket lining residues			
8	A78S, I92M	-	4.0 $\pm$ 1.0	N-cadherin with E-like Trp2 pocket lining residues			
9	D27N, N90D	-	0.64 $\pm$ 0.15	N-cadherin with E-like polar residues around Trp2 pocket			
10	A78S, I92M, D27N, N90D	-	3.4 $\pm$ 1.94	N-cadherin with both E-like polar residues and E-like Trp2 pocket lining residues			
11	W2A	916 $\pm$ 47 ²	26.9 $\pm$ 3.1	Removes swapping Trp2			
12	N-terminal DW-deletion	662 $\pm$ 28.5 ²	42.5 $\pm$ 2.0	Removes swapping N-terminal residues Asp1 and Trp2			
13	N-terminal AA-extension	811 $\pm$ 97 ²	47.2 $\pm$ 2.1	N-terminal addition of two alanines, preventing salt bridge between NH ₃ -term. and residue 89, and impairing swapping			
14	E89A	293 $\pm$ 11 ²	3.0 $\pm$ 0.01	Prevents salt bridge between NH ₃ -term. and residue 89 and impairs swapping			
15	AA-insertion between res. 2 and 3	195 $\pm$ 8.6 ³	3.4 $\pm$ 1.7	Lengthens the A-strand, and decreases swapping driving strain in the monomer			
16	W2F	246.5 $\pm$ 2.1 ³	3.1 $\pm$ 1.7	Decreases swapping driving A-strand strain in the monomer			
17	W2F R14E	n.d.	381 $\pm$ 39	W2F mutation combined with X-incompetent mutation			

^asee methods for measurement details. ^bSee Figure 1C for an illustrative map of these mutations. ^cValue based on a single measurement, ¹Katsamba et al, PNAS 2009, ²Harrison et al, NSMB 2010, ³Vendome et al, NSMB 2011

Table 2 : Additional buried surface area in the swapped dimer relative to the closed monomers.

	N-cad WT	P-cad WT	E-cad WT	C-cad WT	N-cad A78S I92M
Total ($\text{\AA}^2$)	966	1045	989	1008	928
Hydrophobic residues ($\text{\AA}^2$)	746	660	603	628	622

Figures

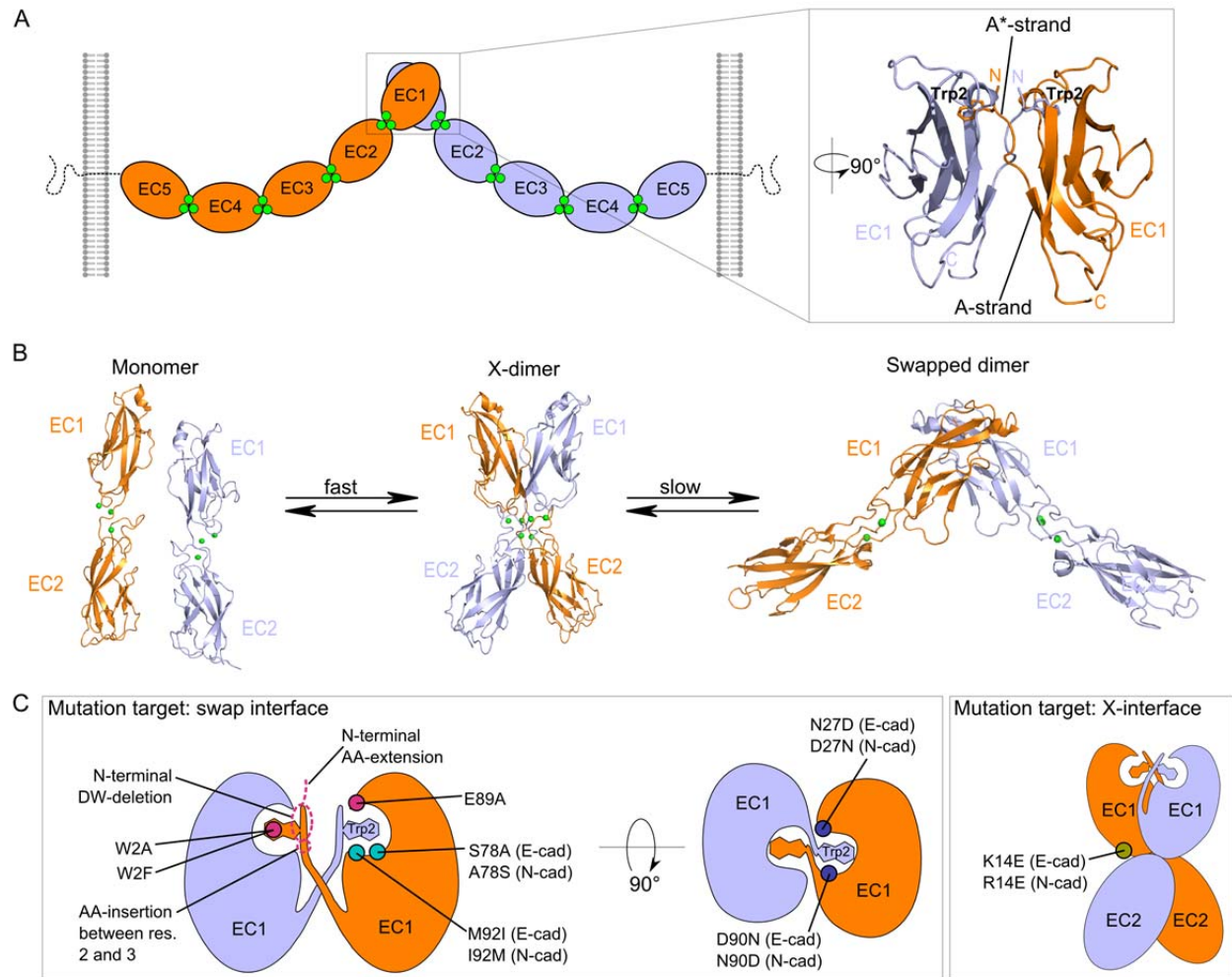


Figure 1. *Trans*-dimerization of type I classical cadherins. Protomers emerging from apposed cells are shown in blue and orange, and calcium ions are shown as green spheres. (A) Schematic illustration of type I classical cadherin *trans* dimer on cell membranes. Extracellular regions dimerize through an interface located in their EC1 domain in which the N-terminal β -strands are swapped, and Trp2 from each protomer is docked in its partner's hydrophobic pocket (expanded view, PDB ID: 2QVF). The A* strand, which consists of the first three residues in the sequence and swaps during dimerization, and the A strand, which consists of the last four residues (7-10) in the first β -strand, are indicated. (B) Reaction scheme showing the X-dimer acting as a kinetic intermediate during the formation of the strand-swapped dimer, in E-cadherin. In the X-dimer (PDB ID: 1FF5), Trp2 is docked in its own protomer's hydrophobic pocket. (C) Schematic illustration of positions of key mutations investigated in this work. They are present on both protomers, but for clarity are shown on one protomer only. The X-dimer-incompetent mutation is indicated in olive (left box); mutations that directly disrupt the strand-swap interface are indicated in pink, those that line the floor of the tryptophan pocket in teal, and those that affect the electrostatic potential at the swap interface are in dark blue (right box).

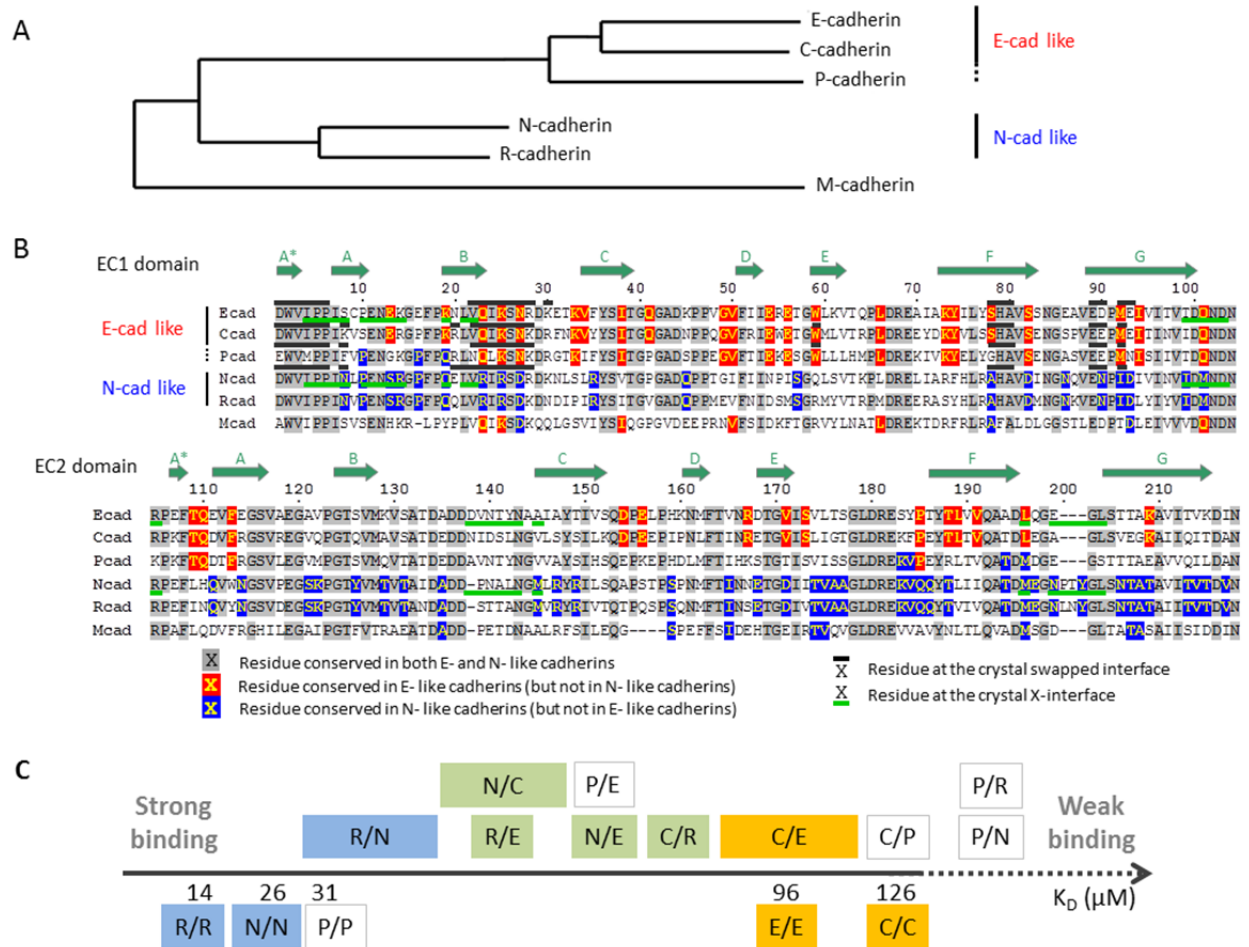


Figure 2. Sequence comparison of type I classical cadherins and scale of relative binding affinities. (A) Phylogenetic tree (obtained with the PhyML method, see material and methods) of classical cadherins based on the EC1 domain sequences of mouse E-, P-, N-, R- and M-cadherin, and *Xenopus* C-cadherin. (B) Sequence alignment of EC1 and EC2 domains. Subtype-specific residues are highlighted in red and blue, and conserved residues in gray. The swapped- and X-dimer interfaces are marked by black (above sequence) and green (below sequence) bars respectively, when crystal structures are available. Positions of the beta-strands are indicated by the arrows above the sequence. (C) Scale of homophilic and heterophilic binding affinities. Homophilic binding affinities, determined by AUC, are given as K_D below the axis. Heterophilic interactions are indicated above the axis, and their specific order along the axis indicates their relative binding affinities, determined by SPR. In the case of the R/N, N/C, and C/E interactions, the relative order was not possible to fully determine, and larger boxes indicate the range of possible affinities. The interactions are colored according to whether they are between among E-like cadherins (in orange), among N-like cadherins (in blue), or across subtypes (in green). Interactions involving P-cadherin are not colored.

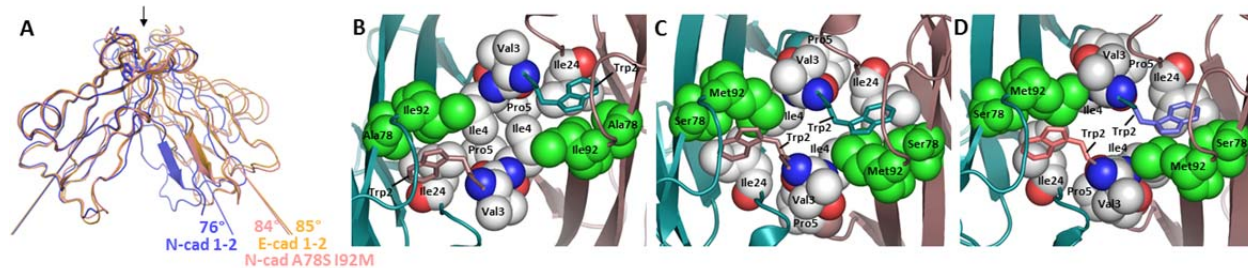


Figure 3. Structural importance of the Trp2 pocket lining residues 78 and 92. (A) Ribbon diagram representation of superposed type I classical cadherins' strand-swapped dimers. EC1 domains of the crystal structure of E-cadherin (2QVF) is shown in orange, N-cadherin wild type (2QVI) in blue, and the N-cadherin double mutant A78S I92M in salmon. For all dimer structures, only the protomer in the foreground has been superposed. The A-strand of the other protomer highlights the difference of relative orientation of the two protomers within each swapped dimer. The long axis of each EC1 domain used to calculate dimer angles (see methods) is represented. Supplementary Figure 3 shows a similar superposition including other available type I cadherin swapped dimer crystal structures. (B-D) Detailed view of side chain packing at and around the hydrophobic binding pocket of Trp2 for wild type N-cadherin (B), wild type E-cadherin (C) and N-cadherin A78S I92M double mutant (D). In each panel the swapped interface is viewed from the top, as indicated by the eye above panel A. The same crystal structures as in panel A have been used. The side chains of residues lining the Trp2 pocket or part of the hydrophobic cluster around it are shown in Van der Waals sphere representation, except for the two Trp2 that are shown in stick representation for clarity. The only subtype-specific residues, 78 and 92, are shown in green. Note that the color coding is different than in panel A as in each dimer one protomer is represented in dark salmon and the other one in cyan.

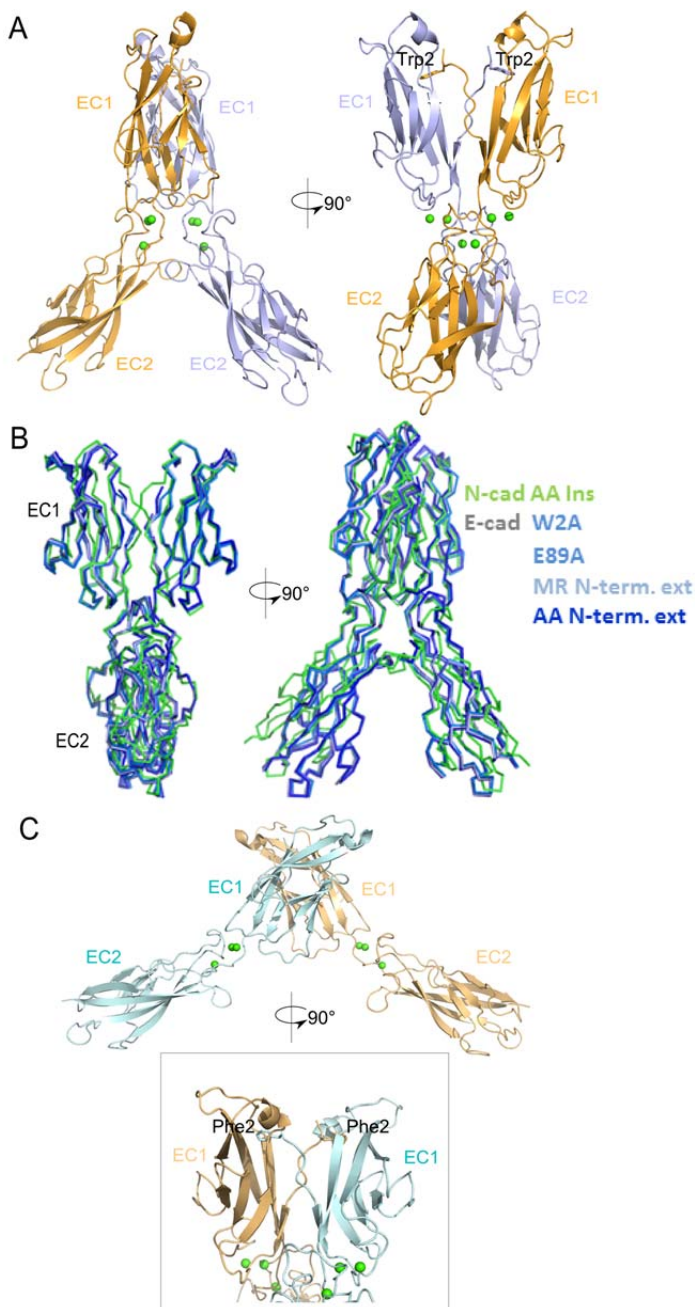


Figure 4. Structure of N-cadherin W2F and N-cadherin AA-insertion mutants. (A) Structure of EC1-2 dimer of N-cadherin AA-insertion mutant. The two protomers are shown in light blue and bright orange, with the exchanged Trp2 side chain shown. Calcium ions are shown as green spheres. (B) Ribbon representation showing the superposition of the N-cadherin AA-insertion mutant (green) and known E-cadherin X-dimer mutants (shades of blue): E-cadherin W2A (3LNH), E89A (3LNI), MR N-terminal extension (1EDH), and AA N-terminal extension (3LNG). (C) Structure of EC1-2 dimer of N-cadherin W2F mutant. Protomers are shown in pale cyan and light orange, calcium ions in green spheres. A close-up of the swapped interface, with the exchanged Phe2 side chains, is shown.

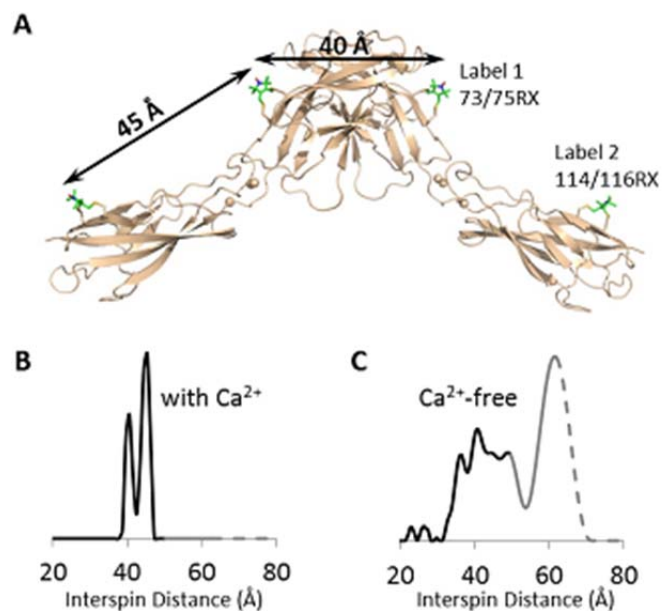


Figure 5. DEER distance distributions analysis of spin-labeled E-cadherin EC1-2 in the presence and absence of calcium. (A) Sites of bifunctional RX side chains (in stick representation) are shown on E-cadherin structure (wheat cartoon, PDB ID: 2QVF). The interspin distance between 73/75RX and 114/116RX on the same protomer is estimated by modeling to be 45 Å, and the distance between 73/75RX on one protomer and its dimer partner is approximately 40 Å. (B-C) Distance probability distribution of E-cadherin EC1-2 in the presence (B) and absence (C) of calcium. Probabilities at distances longer than the upper limit of accurate peak width determination are shown in gray; those at distances longer than the upper limit of accurate peak position determination are shown as dashes.

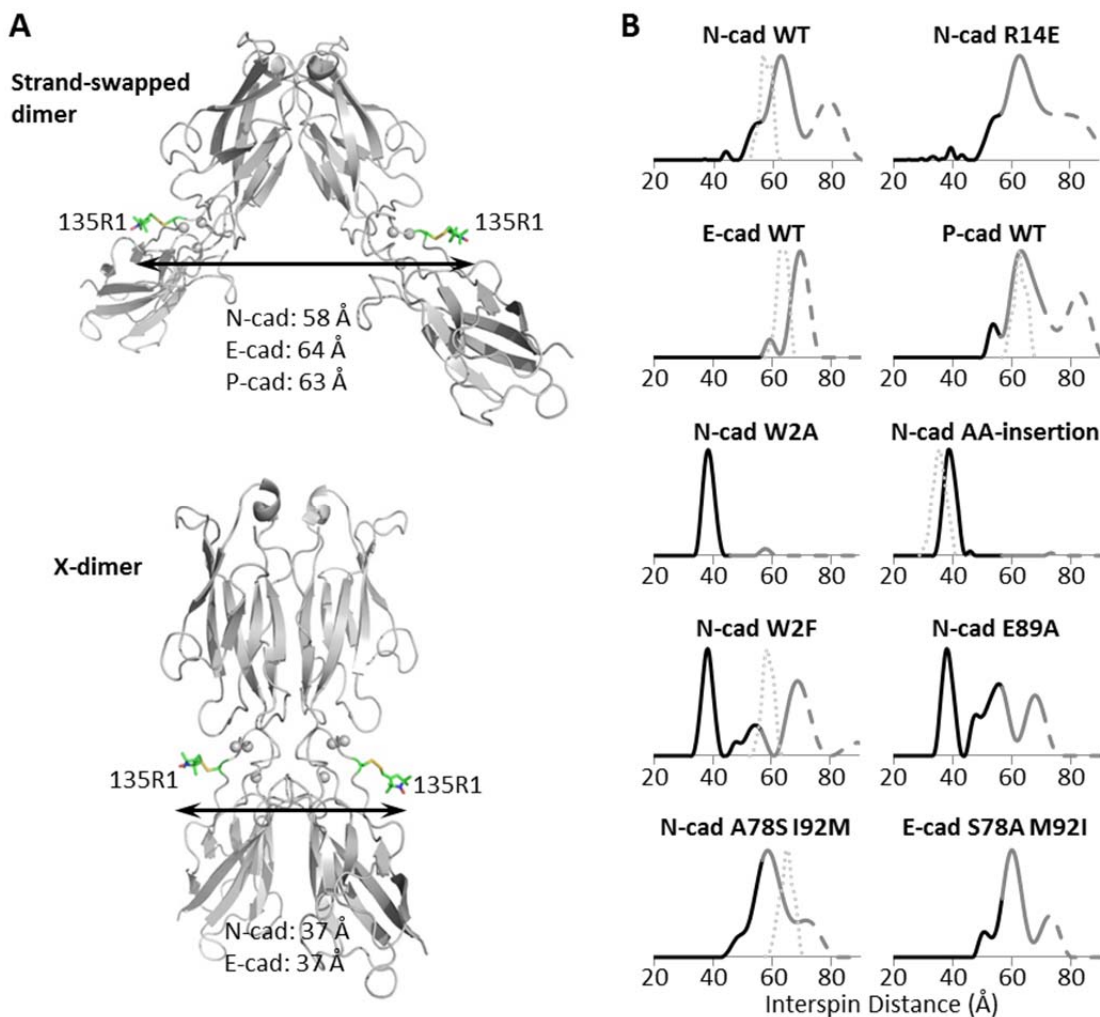


Figure 6. DEER distance distributions of dimers for spin-labeled N-, E-, P-cadherins and N-, E-mutants. (A) Site of spin labeling (R1 in stick representation) is shown on N-cadherin structures (gray cartoon, showing wild-type (PDB ID: 2QVI) and the AA-insertion mutant (PDB ID: 4NUP), with calcium ions as gray spheres). The average interspin distance in the N-, E-, and P-cadherin swapped dimer (top) is estimated to be 58, 64, and 63 Å, respectively, and in the X-dimer (bottom), 37 Å for both N- and E-cadherin. (B) Distance probability distributions of wild-type N-, E-, and P-cadherin EC1-2, and mutant N- and E-cadherin EC1-2. Thin dotted lines represent predicted interspin distance distributions based on crystal structures, if available, and accounting for spin label rotamer conformations (see Methods). All constructs include the 135C mutation for probe labeling; E-cadherin constructs include the C9S mutation of the native cysteine. Probabilities at distances longer than the upper limit of accurate peak width determination are shown in gray; those at distances longer than the upper limit of accurate peak position determination are shown as dashes.

Supplementary material

Supplementary Table 1: Data collection and refinement statistics

	P-cad wild-type	N-cad A78S I92M	N-cad W2F	N-cad AA insertion
Data Collection				
Space group	P2 ₁ 2 ₁ 2	P2 ₁	C2	P2 ₁ 2 ₁ 2
Cell dimensions: a,b,c (Å)	123.1, 188.7, 53.2	59.8, 221.2, 72.2	116.6, 86.2, 46.7	175.5, 65.8, 102.5
α, β, γ (°)	90, 90, 90	90, 103.9, 90	90, 98.5, 90	90, 90, 90
Molecules per asymmetric unit	4	4	1	3
Resolution limit (Å)	40-3.2	20-3.2	20-2.1	20-2.7
Unique reflections	21058	25726	25187	33781
Redundancy (Highest resolution shell)	6.3 (6.1)	3.5 (1.9)	3.4 (3.0)	5.3 (3.7)
Completeness % (Highest resolution shell)	99.3 (94.2)	94.7 (68.2)	97.7 (85.7)	99.7 (96.9)
Average I/ σ (I) (Highest resolution shell)	11.6 (3.1)	9.5 (2.1)	34.3 (3.0)	13.6 (2.0)
Rmerge (%) (Highest resolution shell)	11.6 (58.9)	12.0 (44.5)	11.2 (46.1)	11.6 (44.4)
Refinement				
Resolution limit (Å)	20-3.2	20-3.2	20-2.1	20-2.7
Rwork (%)	22.8	21.9	21.1	17.2
Rfree (%)	26.9	25.3	25.3	22.9
Rmsd bonds (Å)	0.01	0.006	0.008	0.009
Rmsd angles (°)	0.9	1.07	1.08	1.26
Protein atoms	6552	6668	1659	5042
Ligand/ion atoms	23	12	3	10
Water molecules	76	0	159	576
Average B (Å ²) protein atoms	71.4	68.5	53.5	23.0
Average B (Å ²) ligand/ion atoms	63.8	60.7	48.9	26.4
Average B (Å ²) water molecules	24.1	N/A	57.6	23.9

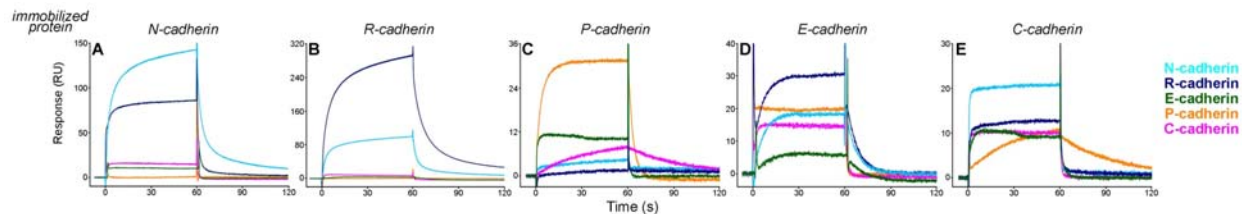
Supplementary Table 2:

Total and hydrophobic surface area buried at the swapped interfaces as determined from crystal structures of N-, E-, C-, P-cadherins and the N-cadherin mutant A78S I92M. The EC1-EC1 dimer angle is indicated for each crystal structure.

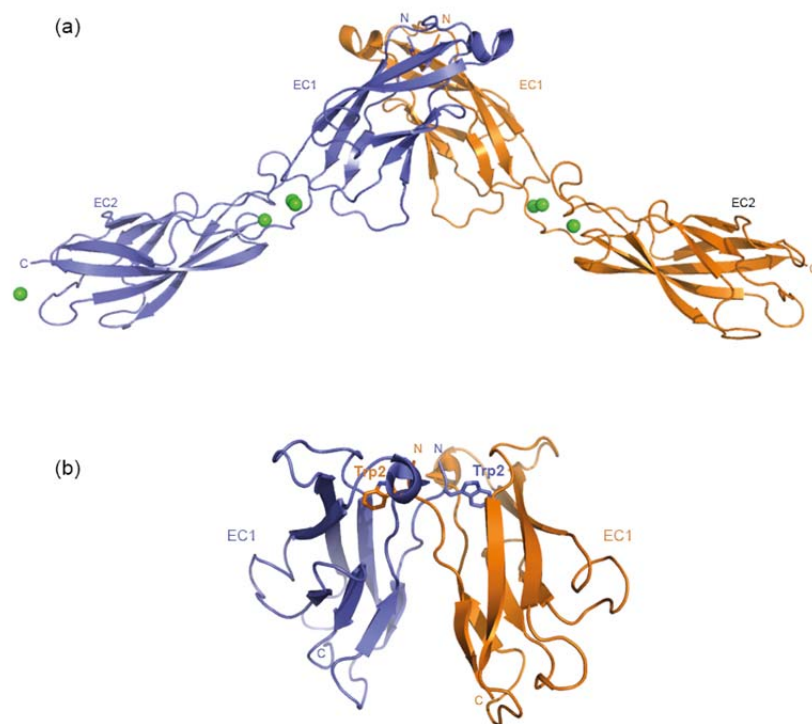
	Residues 78 and 92	Total BSA (Å²)	Hydrophobic BSA (Å², by residue)	EC1-EC1 dimer angle (degree)
N-cad WT EC1-2 (2QVI)	A78, I92	1839	1202	75.7
N-cad WT EC1-5 (3Q2W)	A78, I92	1900	1218	73.4
N-cad W2F (4NUQ)	A78, I92	1796	1149	77.6
N-cad A78S I92M (4NUM <i>form 1</i>)	S78, M92	1747	998	84.2
N-cad A78S I92M (4NUM <i>form 2</i>)	S78, M92	1751	1027	83.9
E-cad WT mouse (2QVF)	S78, M92	1834	1008	85.3
E-cad WT mouse EC1-5 (3Q2V)	S78, M92	1725	962	89.2
E-cad WT human (2O72)	S78, M92	1800	1031	88.1
C-cad WT (1L3W)	S78, M92	1840	970	88.7
P-cad WT (4NQQ)	G78, M92	1850	996	83.4

Supplementary Table 3: Dissociation constants (K_D) for wild type E- and N-cadherins and their pocket mutants. Data are given as mean $\pm$ s.d.

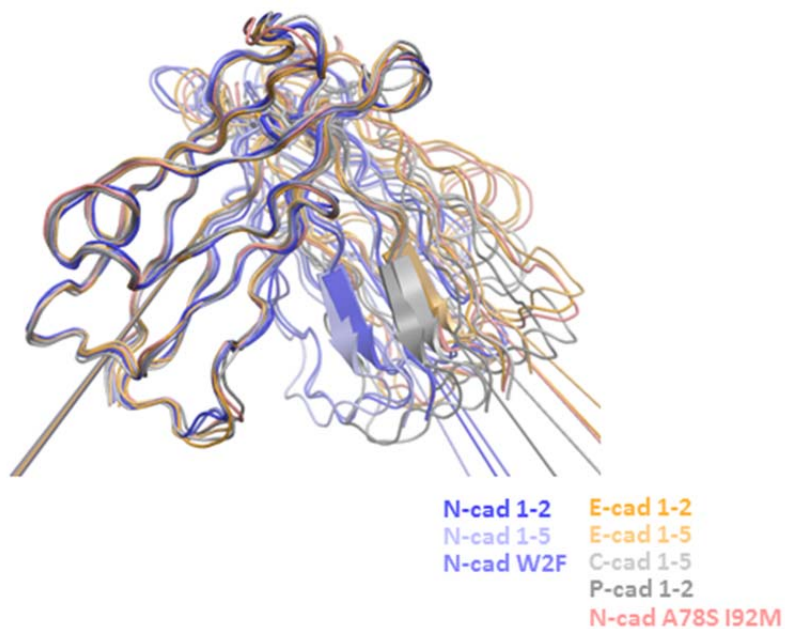
Mutation	K_D (μM)		Mutant description
	E-cadherin	N-cadherin	
M92I	54.6 $\pm$ 3.8	-	E-cadherin with N-like Trp2 pocket lining residue 92
S78A	85.7 $\pm$ 3.9	-	E-cadherin with N-like Trp2 pocket lining residue 78
I92M	-	2.6 $\pm$ 0.2	N-cadherin with E-like Trp2 pocket lining residue 92
A78S	-	34.4 $\pm$ 2.0	N-cadherin with E-like Trp2 pocket lining residue 78



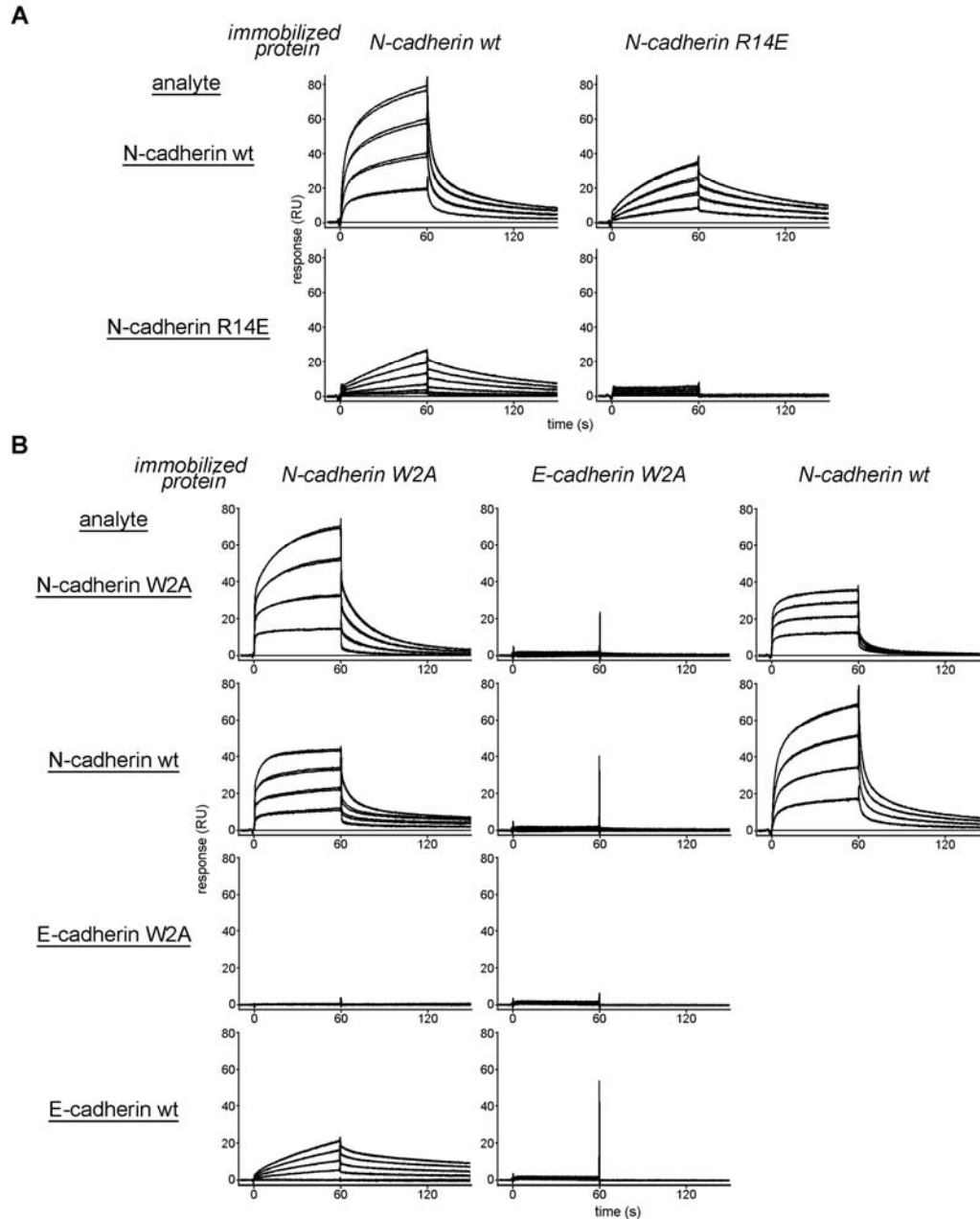
Supplementary Figure 1. SPR analysis of the heterophilic binding of type I cadherins. Each of the five type I cadherins, C-, E-, N-, P- and R-cadherin analytes was injected at 12 μ M monomer concentration over individual sensor chip surfaces immobilized with each of the five type I cadherin at 70 μ M monomer concentration. The responses are color-coded as indicated by the figure legend. The immobilized molecule for each panel proteins is shown in italics at the top.



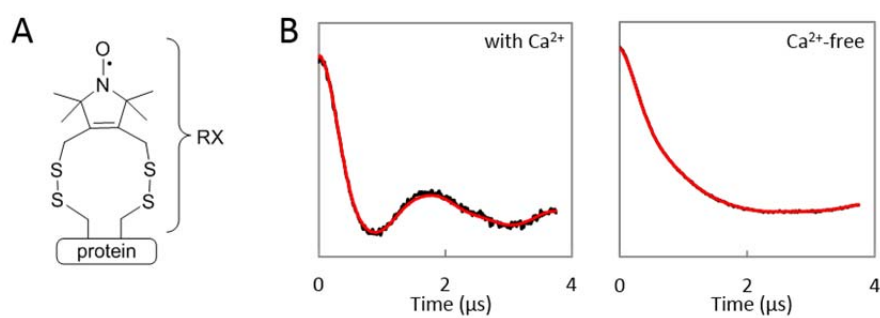
Supplementary Figure 2. Crystal structure of P-cadherin adhesive fragment EC1-2 reveals an adhesive strand swapped dimer. (A) Strand-swapped dimer formed by wild-type P-cadherin observed in the crystal structure. One protomer shown in blue ribbon presentation; the binding partner shown in orange; green spheres represent calcium (II) ions. (B) Close-up of EC1 domains showing exchanged A*-strands and docking of Trp2 residues (stick representation). Dimer formed between chains A and B (pdb 4NQQ) is depicted in the figure. Chain C and D also adopt a strand swapped dimer configuration, however, EC1 of chain D is poorly ordered.



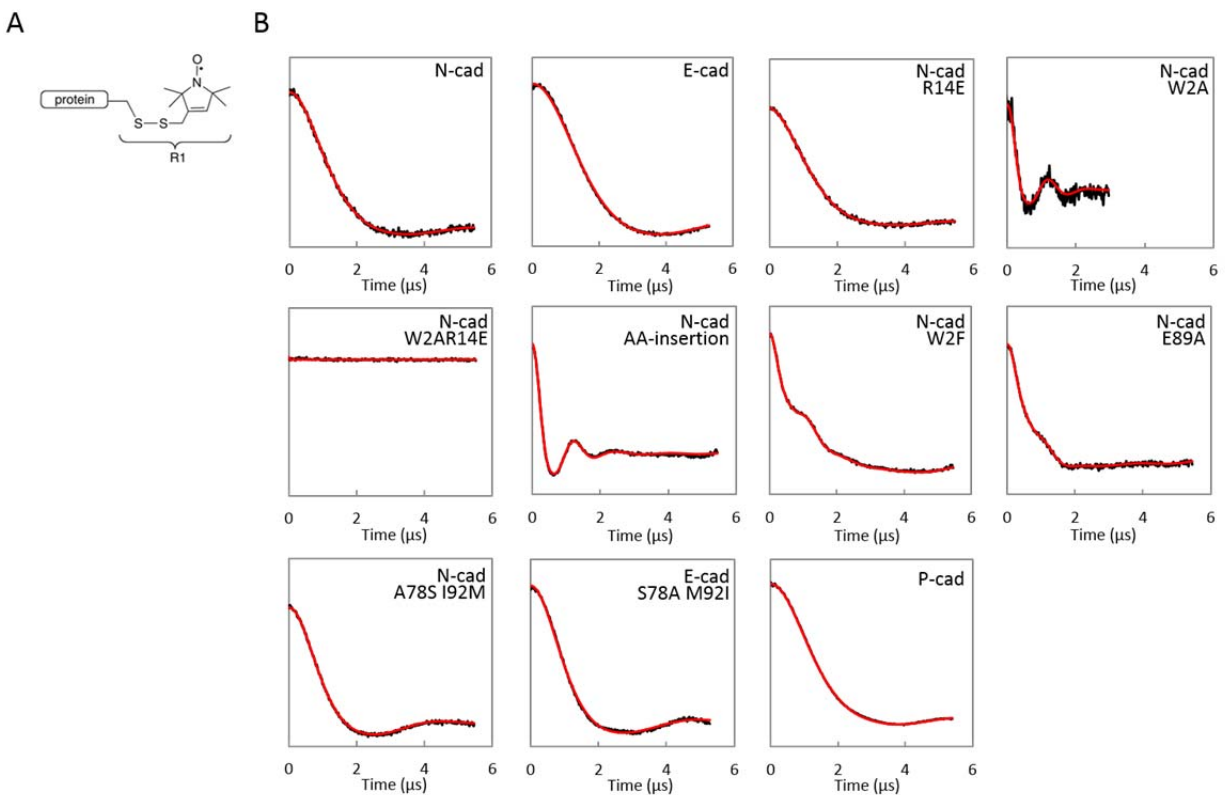
Supplementary Figure 3. Ribbon diagram representation of superposed type I classical cadherins' strand-swapped dimers. EC1 domains of the crystal structure of E-cadherin EC1-2 (2QVF), E-cadherin EC1-5 (3Q2V), C-cadherin EC1-5 (1L3W), P-cadherin, N-cadherin EC1-2 (2QVI), N-cadherin EC1-5 (3Q2W) and the N-cadherin double mutant A78S I92M are shown. The color code is indicated on the figure. For all dimer structures, only the protomer in the foreground has been superposed. The A-strand of the other protomer highlights the difference of relative orientation of the two protomers within each swapped dimer. The A-strand of the other protomer highlights the difference of relative orientation of the two protomers within each swapped dimer. The long axis of each EC1 domain used to calculate dimer angles (see methods) are represented.



Supplementary Figure 4. SPR analysis of W2A strand-swap mutant and R14E X-dimer mutants of N-cadherin. For each panel, the immobilized proteins are indicated in italics at the top and the injected analytes are shown underlined at the left-hand side. All panels are depicted on the same scale. Proteins were immobilized at concentrations corresponding to 100 μ M monomer and analyte binding was tested at concentrations corresponding to 12, 9, 6 and 3 μ M monomer. (A) Binding of N-cadherin wild type and its R14E X-dimer mutant, to surfaces immobilized with N-cadherin wild-type and N-cadherin R14E. (B) N-cadherin W2A and E-cadherin W2A strand-swap mutants were tested for binding over surfaces immobilized with N-cadherin W2A, E-cadherin W2A and N-cadherin wild type proteins. The binding of the wild type N- and E-cadherins was included as a control.



Supplementary Figure 5. DEER analysis of RX-labeled E-cadherin. (A) Chemical structure of the RX side chains. (B) Background-subtracted dipolar evolutions of E-cadherin 73/75RX 114/116RX in the presence (left) and absence (right) of Ca²⁺. Both the experimental form factor (black trace) and the best fit (red trace) from the program LongDistances are shown.



Supplementary Figure 6. DEER analysis of R1-labeled E-, N-, P-cadherin and mutants. (A) Chemical structure of the R1 side chain. (B) Background-subtracted dipolar evolutions of E-, N-, P-cadherin and indicated mutants, all labeled with R1 at position 135. Constructs for E-cadherin include an additional C9S mutation to remove the native cysteine. Both the experimental form factor (black trace) and the best fit (red trace) from the program LongDistances are shown.

References

1. Takeichi M (1990) Cadherins: a molecular family important in selective cell-cell adhesion. *Annual review of biochemistry* 59:237-252.
2. Takeichi M (1995) Morphogenetic roles of classic cadherins. *Current opinion in cell biology* 7(5):619-627.
3. Halbleib JM & Nelson WJ (2006) Cadherins in development: cell adhesion, sorting, and tissue morphogenesis. *Genes & development* 20(23):3199-3214.
4. Gumbiner BM (2005) Regulation of cadherin-mediated adhesion in morphogenesis. *Nat Rev Mol Cell Biol* 6(8):622-634.
5. Suzuki SC & Takeichi M (2008) Cadherins in neuronal morphogenesis and function. *Development, growth & differentiation* 50 Suppl 1:S119-130.
6. Price SR, De Marco Garcia NV, Ranscht B, & Jessell TM (2002) Regulation of motor neuron pool sorting by differential expression of type II cadherins. *Cell* 109(2):205-216.
7. Duguay D, Foty RA, & Steinberg MS (2003) Cadherin-mediated cell adhesion and tissue segregation: qualitative and quantitative determinants. *Developmental biology* 253(2):309-323.
8. Foty RA & Steinberg MS (2005) The differential adhesion hypothesis: a direct evaluation. *Developmental biology* 278(1):255-263.
9. Katsamba P, *et al.* (2009) Linking molecular affinity and cellular specificity in cadherin-mediated adhesion. *Proc Natl Acad Sci U S A* 106(28):11594-11599.
10. Boggon TJ, *et al.* (2002) C-cadherin ectodomain structure and implications for cell adhesion mechanisms. *Science* 296(5571):1308-1313.
11. Patel SD, *et al.* (2006) Type II cadherin ectodomain structures: implications for classical cadherin specificity. *Cell* 124(6):1255-1268.
12. Parisini E, Higgins JM, Liu JH, Brenner MB, & Wang JH (2007) The crystal structure of human E-cadherin domains 1 and 2, and comparison with other cadherins in the context of adhesion mechanism. *J Mol Biol* 373(2):401-411.
13. Harrison OJ, *et al.* (2011) The extracellular architecture of adherens junctions revealed by crystal structures of type I cadherins. *Structure* 19(2):244-256.
14. Nagar B, Overduin M, Ikura M, & Rini JM (1996) Structural basis of calcium-induced E-cadherin rigidification and dimerization. *Nature* 380(6572):360-364.
15. Pertz O, *et al.* (1999) A new crystal structure, Ca²⁺ dependence and mutational analysis reveal molecular details of E-cadherin homoassociation. *EMBO J* 18(7):1738-1747.
16. Harrison OJ, *et al.* (2010) Two-step adhesive binding by classical cadherins. *Nat Struct Mol Biol* 17(3):348-357.
17. Li Y, *et al.* (2013) Mechanism of E-cadherin dimerization probed by NMR relaxation dispersion. *Proc Natl Acad Sci U S A* 110(41):16462-16467.
18. Hong S, Troyanovsky RB, & Troyanovsky SM (2010) Spontaneous assembly and active disassembly balance adherens junction homeostasis. *Proc Natl Acad Sci U S A* 107(8):3528-3533.
19. Harrison OJ, Corps EM, & Kilshaw PJ (2005) Cadherin adhesion depends on a salt bridge at the N-terminus. *Journal of cell science* 118(Pt 18):4123-4130.
20. Tamura K, Shan WS, Hendrickson WA, Colman DR, & Shapiro L (1998) Structure-function analysis of cell adhesion by neural (N-) cadherin. *Neuron* 20(6):1153-1163.
21. Troyanovsky RB, Sokolov E, & Troyanovsky SM (2003) Adhesive and lateral E-cadherin dimers are mediated by the same interface. *Molecular and cellular biology* 23(22):7965-7972.
22. Shan WS, Koch A, Murray J, Colman DR, & Shapiro L (1999) The adhesive binding site of cadherins revisited. *Biophysical chemistry* 82(2-3):157-163.
23. Niessen CM & Gumbiner BM (2002) Cadherin-mediated cell sorting not determined by binding or adhesion specificity. *The Journal of cell biology* 156(2):389-399.

24. Shan WS, *et al.* (2000) Functional cis-heterodimers of N- and R-cadherins. *The Journal of cell biology* 148(3):579-590.
25. Chen C, Posy S, Ben-Shaul A, Shapiro L, & Honig B (2005) Specificity of cell-cell adhesion by classical cadherins: Critical role for low-affinity dimerization through beta-strand swapping. *Proceedings of the National Academy of Sciences of the United States of America* 102(24):8531-8536.
26. Vendome J, *et al.* (2011) Molecular design principles underlying beta-strand swapping in the adhesive dimerization of cadherins. *Nat Struct Mol Biol* 18(6):693-700.
27. Vunnam N & Pedigo S (2011) Calcium-induced strain in the monomer promotes dimerization in neural cadherin. *Biochemistry* 50(39):8437-8444.
28. Hulpiau P & van Roy F (2009) Molecular evolution of the cadherin superfamily. *Int J Biochem Cell Biol* 41(2):349-369.
29. Ciatto C, *et al.* (2010) T-cadherin structures reveal a novel adhesive binding mechanism. *Nat Struct Mol Biol* 17(3):339-347.
30. Jeschke G (2012) DEER distance measurements on proteins. *Annu Rev Phys Chem* 63:419-446.
31. Fleissner MR, *et al.* (2011) Structure and dynamics of a conformationally constrained nitroxide side chain and applications in EPR spectroscopy. *Proc Natl Acad Sci U S A* 108(39):16241-16246.
32. Fleissner MR, Cascio D, & Hubbell WL (2009) Structural origin of weakly ordered nitroxide motion in spin-labeled proteins. *Protein Sci* 18(5):893-908.
33. Wright PE & Dyson HJ (2009) Linking folding and binding. *Curr Opin Struct Biol* 19(1):31-38.
34. Wu Y, Vendome J, Shapiro L, Ben-Shaul A, & Honig B (2011) Transforming binding affinities from three dimensions to two with application to cadherin clustering. *Nature* 475(7357):510-513.
35. Harrison OJ, *et al.* (2012) Nectin ectodomain structures reveal a canonical adhesive interface. *Nat Struct Mol Biol* 19(9):906-915.
36. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics* 5:113.
37. Guindon S, Lethiec F, Duroux P, & Gascuel O (2005) PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res* 33(Web Server issue):W557-559.
38. Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Macromolecular Crystallography, Pt A* 276:307-326.
39. McCoy AJ, *et al.* (2007) Phaser crystallographic software. *Journal of applied crystallography* 40(Pt 4):658-674.
40. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D* 60:2126-2132.
41. Adams PD, *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography* 66(Pt 2):213-221.
42. Murshudov GN, Vagin AA, & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta crystallographica. Section D, Biological crystallography* 53(Pt 3):240-255.
43. Morin A, *et al.* (2013) Collaboration gets the most out of software. *eLife* 2:e01456.
44. Sridharan S, Nicholls A, & Honig B (1992) A New Vertex Algorithm to Calculate Solvent Accessible Surface-Areas. *Faseb J* 6(1):A174-A174.
45. Foote J & Raman A (2000) A relation between the principal axes of inertia and ligand binding. *Proc Natl Acad Sci U S A* 97(3):978-983.
46. Brooks BR, *et al.* (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* 30(10):1545-1614.

47. Petrey D, *et al.* (2003) Using multiple structure alignments, fast model building, and energetic analysis in fold recognition and homology modeling. *Proteins* 53 Suppl 6:430-435.
48. Grassetti DR & Murray JF, Jr. (1967) Determination of sulfhydryl groups with 2,2'- or 4,4'-dithiodipyridine. *Arch Biochem Biophys* 119(1):41-49.
49. Hatmal MM, *et al.* (2012) Computer modeling of nitroxide spin labels on proteins. *Biopolymers* 97(1):35-44.
50. Altenbach C, Kusnetzow AK, Ernst OP, Hofmann KP, & Hubbell WL (2008) High-resolution distance mapping in rhodopsin reveals the pattern of helix movement due to activation. *Proc Natl Acad Sci U S A* 105(21):7439-7444.
51. Lopez CJ, Yang Z, Altenbach C, & Hubbell WL (2013) Conformational selection and adaptation to ligand binding in T4 lysozyme cavity mutants. *Proc Natl Acad Sci U S A* 110(46):E4306-4315.