Dimerization mechanism and structural features of human LI-cadherin

Anna Yui[1], Jose M. M. Caaveiro[1,2], Daisuke Kuroda[1,3], Makoto Nakakido[1], Satoru Nagatoishi[4], Shuichiro Goda[5], Takahiro Maruno[6], Susumu Uchiyama[6] and Kouhei Tsumoto[1,4,7]*

[1]Department of Bioengineering, Graduate School of Engineering, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

[2]Department of Global Healthcare, Graduate School of Pharmaceutical Sciences, Kyushu University, 3-1-1, Maidashi, Higashi-ku, Fukuoka-shi, Fukuoka 812-8582, Japan

[3]Medical Device Development and Regulation Research Center, School of Engineering, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

[4]Institute of Medical Science, The University of Tokyo, 4-6-1, Shirokanedai, Minato-ku, Tokyo, 108-8639, Japan

[5]Graduate School of Science and Engineering, Soka University, 1-236, Tangi-cho, Hachioji-shi, Tokyo 192-8577 Japan

[6]Department of Biotechnology, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita-shi, Osaka 565-0871, Japan

[7]Department of Chemistry and Biotechnology, School of Engineering, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

*Corresponding author: Kouhei Tsumoto

E-mail: tsumoto@bioeng.t.u-tokyo.ac.jp;

Running title: Dimerization mechanism of LI-cadherin

Keywords: Cadherin, dimerization, cell adhesion, protein chemistry, crystal structure, small-angle X-ray scattering (SAXS), analytical ultracentrifugation, molecular dynamics

Abstract

LI-cadherin is a member of cadherin superfamily which is a Ca^{2+}-dependent cell adhesion protein. Its expression is observed on various types of cells in the human body such as normal small intestine and colon cells, and gastric cancer cells. Because its expression is not observed on normal gastric cells, LI-cadherin is a promising target for gastric cancer imaging. However, since the cell adhesion mechanism of LI-cadherin has remained unknown, rational design of therapeutics targeting this cadherin has been complicated. Here, we have studied the homodimerization mechanism of LI-cadherin. We report the crystal structure of the LI-cadherin EC1-4 homodimer. The EC1-4 homodimer exhibited a unique architecture different from that of other cadherins reported so far. The crystal structure also revealed that LI-cadherin possesses a noncanonical calcium ion-free linker between EC2 and EC3. We also show that LI-cadherin EC1-2 and EC3-4 have different characteristics to that of the EC1-2 domains of classical cadherins despite the sequence similarity among them. Various biochemical techniques, molecular dynamics (MD) simulations and small angle X-ray scattering (SAXS) were employed to elucidate the mechanism of homodimerization. We expect these findings will advance the generation of therapeutic molecules targeting LI-cadherin and to establish the specific role of LI-cadherin on cancer cells.

Introduction

Cadherins are a family of glycoproteins responsible for calcium ion-dependent cell adhesion (1). There are more than 100 types of cadherins in humans and many of them are responsible not only for cell adhesion but also involved in tumorigenesis (2). Human liver intestine-cadherin (LI-cadherin) is a nonclassical cadherin comprising seven extracellular domains (3). Previous studies have reported the expression of LI-cadherin on various types of cells, such as normal intestine cells, intestinal metaplasia, colorectal cancer cells and lymph node metastatic gastric cancer cells (4, 5).

Because human LI-cadherin is expressed on gastric cancer cells but not on normal stomach tissues, LI-cadherin has been proposed as a target for imaging metastatic gastric cancer (6). Previous studies have reported that LI-cadherin works not only as a calcium ion-dependent cell adhesion molecule as other cadherins do (7), but also shown that trans-dimerization of LI-cadherin is necessary for water transport in normal intestinal cells (8). Sequence analysis of mouse LI-, E-, N-, and P-cadherins has revealed sequence homology between EC1-2 of LI-cadherin and EC1-2 of E-, N-, and P-cadherins, as well as between EC3-7 of LI-cadherin and EC1-5 of classical cadherins (9). From the sequence similarity and the proposed absence of calcium ion-binding motifs (10, 11) between domains EC2 and EC3, there is speculation that LI-cadherin has evolved from the same five-domain cadherin

precursor as that of classical cadherins (9).

However, LI-cadherin is different from classical cadherins in some points such as the number of extracellular domains and the length and the sequence of the cytoplasmic domain. Classical cadherins possess five extracellular domains whereas LI-cadherin possesses seven (2). Classical cadherins possess a conserved cytoplasmic domain comprising more than 100 amino acids, whereas LI-cadherin possess a short cytoplasmic domain consisting of 20 residues with little or no sequence homology (7, 12).

The characteristics of LI-cadherin at the molecular level, including the homodimerization mechanism, still remains unknown. Homodimerization is the fundamental event in cadherin-mediated cell adhesion as has been shown previously (13, 14). In particular, classical cadherins are reported to form a homodimer mediated by the interaction between their two N-terminal domains (EC1-2) (10, 15).

In this study, we aimed to characterize LI-cadherin at the molecular level as understanding the molecular characteristics of target protein would lead to the rational design of medicine. We have extensively validated LI-cadherin to identify the dimerization architecture of LI-cadherin. Here, we report the crystal structure of human LI-cadherin EC1-4 homodimer. This crystal structure revealed a dimerization architecture different from that of any other cadherin reported so far. It also showed canonical calcium binding motifs between EC1 and EC2 domains, and between EC3 and EC4, but not between EC2 and EC3. By performing various biochemical and computational analysis based on this crystal structure, we interpreted the characteristics of LI-cadherin molecule. In particular, small-angle X-ray scattering (SAXS) showed that LI-cadherin EC1-2 and EC3-4 behaved differently to domains EC1-2 of classical cadherins, even though the sequence similarity between them. Our study revealed possible architectures of LI-cadherin homodimers at the cell surface and suggested the differential role of the two additional domains at the N-terminus compared with classical cadherins.

Results

Crystal structure analysis of EC1-4 homodimer

In order to reveal the structure of the homodimer of human LI-cadherin we solved the X-ray crystal structure of EC1-4 (Table S1) at 2.7 Å resolution (Fig. 1A and Table 1). Each extracellular cadherin (EC) domain was composed of the typical seven β-strands seen in classical cadherins, and three calcium ions bound to each of the linkers connecting EC1 and EC2, and EC3 and EC4 (Fig. 1A). We also observed four N-glycans and two N-glycans bound to chain A and B, respectively, as could be predicted from amino acid sequence. We could not resolve the entire length of these N-glycans because of their high flexibility. From the

portion resolved, all N-glycans seem to face the opposite side of the dimer interface. Two unique characteristics were observed in the crystal structure of LI-cadherin: (i) the architecture of the homodimer and (ii) the existence of a calcium-free linker between EC2 and EC3.

The EC1-4 region of LI-cadherin assembled as an antiparallel homodimer in a conformation different from that of other cadherins, such as classical cadherins, which exhibit two step binding mode (15) and to that of protocadherin γB3, which forms an antiparallel homodimer (14) but with distinct characteristics to that of LI-cadherin EC1-4. A previous study had suggested that LI-cadherin lacks a calcium-binding motif between EC2 and EC3 (9) and our crystal structure confirmed that hypothesis experimentally. This result supports a domain-duplicating event during the evolution of LI-cadherin, facilitating a distinct dimerization strategy by possessing two extra domains in the N-terminal region.

The dissociation constant (K_D) of the EC1-4 homodimer determined by sedimentation velocity analytical ultracentrifugation (AUC) was 39.8 μM (Fig. 1B), and that of the EC1-5 homodimer measured by the same instrument was 22.8 μM (Fig. 1C). The slight increase of the affinity suggested some contact between EC1 and EC5, as can be predicted from the arrangement of EC1 of one chain and EC4 of the other chain in the crystal structure, although this interaction does not seem strong. Interestingly, we observed minor peaks corresponding to the tetramer in AUC measurements of EC1-5 (Fig. S1A). We could not determine the K_D value of tetramerization because the detected volume was very small. These results suggested the existence of weak interactions which could not be detected by other methods. Following the analysis of the crystal structure we employed biochemical and computational methods to further investigate the interactions observed in the crystal structure, and the nature of the calcium-free linker between EC2 and EC3.

Interaction analysis of EC1-4 homodimer

We analyzed the interaction between two EC1-4 molecules using the PISA server (Table S2) (16). The interaction was mostly mediated by EC2 of one chain of LI-cadherin and EC4 of the other chain, engaging in hydrogen bonds and hydrophobic contacts (Fig. 2A). The dimerization interface area was 1,253.3 Å^2 and the number of hydrogen bonds in which the distance between heavy atoms were less than 3.5 Å was seven.

Based on the analysis of these interactions we conducted site-directed mutagenesis to assess the contribution of several residues to the dimerization of LI-cadherin. Eleven residues showing a percentage of buried area greater than 50%, or one or more intermolecular hydrogen bonds (distance between heavy atoms < 3.5 Å) were individually mutated to Ala and their potential to form

homodimers was evaluated by SEC-MALS. The observed weight-averaged molecular weight (MW) of each mutant is shown in Table 2. The value of MW determined by SEC-MALS was revealed to be between the MW of the monomer (49.8 kDa) and that for the dimer (99.6 kDa). Because the samples were diluted to around 6 μM during their movement inside the chromatographic column, even in samples injected at 100 μM, the values determined for WT and mutants were closer to the calculated MW of the monomer than that of the dimer. Considering that the K_D of EC1-4 wild type (WT) homodimer determined by AUC was 39.8 μM, the largest fraction of EC1-4 WT should be mostly monomeric at a protein concentration of 6 μM (the protein concentration estimated at the end of the chromatographic column). However, the equilibrium between monomer and dimer may shift the position of the elution peak in a concentration-dependent manner if indeed such dimer was formed. We have qualitatively compared the tendency of LI-cadherin to dimerize by comparing the MW of each mutant at a constant concentration and also by investigating the concentration-dependent shift of their MW.

It was revealed that homodimerization of the mutant F224A of EC1-4 was almost completely inhibited (Fig. 2B and Table 2). The MW of EC1-4 F224A at each concentration was almost the same as that of the monomer (49.8 kDa) and concentration-dependent shift in the elution volume and the increase of MW was largely absent when compared with WT protein. Dimerization of EC1-4 I169A, EC1-4 L171A, EC1-4 V210A, EC1-4 N222A, EC1-4 N371A and EC1-4 Y399A was inhibited to some extent, however, its effect was less notorious than that observed in EC1-4 F224A (Table 2 and Fig. S2).

Contribution of Phe224 to dimerization

F224 belongs to the β6-β7 loop of the EC2 domain. Although F224 does not seem to form extensive specific interactions with the partner molecule of LI-cadherin in our crystal structure (Fig. S3), its buried area upon dimerization was calculated to be 94% by the PISA server, engaging in Van der Waals interactions between Phe224 and other residues of EC1-4. To understand the role of Phe224 in dimerization of LI-cadherin, we conducted molecular dynamics (MD) simulations of EC1-4 WT and EC1-4 F224A in the monomeric states, respectively. After confirming the convergence of the simulations by calculating RMSD values of Cα atoms (Fig. S4A, B, see Experimental Procedures for the details), we first calculated the distance between Cα atoms of the residues 224 and 122. The simulations revealed that Ala224 moves away from the strand that contains Asn122 whereas Phe224 kept closer to Asn122 (Fig. 3A and Movie 1, 2). This movement suggests that the side chain of Phe224 forms intramolecular interaction and is stabilized inside the pocket.

Thermal stability analysis using differential

scanning calorimetry (DSC) revealed that EC1-4 F224A had two unfolding peaks whereas that of EC1-4 WT had a single peak (Fig. 3B). These results showed that a part of EC1-4 F224A molecule was destabilized by the mutation. In combination with the data from MD simulations, we propose that Phe224 contributes to dimerization of LI-cadherin by intermolecular interactions and shape complementarity between the two monomers of LI-cadherin. DSC measurements showed that some of other mutants have lower thermal stability than wild type (Table 2 and Fig. S5). However, because T_{M1} of F224A is lowest among the mutants evaluated, and because other mutants displaying lower T_{M1} than wild type did not exhibit a drastic decrease in homodimer affinity like F224A, we conclude that among the residues evaluated by Ala scanning, F224 was the most critical for the maintenance of homodimer structure and thermal stability.

To investigate if the mutation to Phe224 influences cell adhesion, we established a CHO cell line expressing full-length LI-cadherin WT or the mutant F224A (including the transmembrane and cytoplasmic domains fused to GFP) that we termed EC1-7GFP and EC1-7F224AGFP (Table S1 and Fig. S6). We conducted cell aggregation assays and compared the cell adhesion ability of cells expressing each construct in the presence of calcium with that of cells in the presence of EDTA or with mock cells (non-transfected Flp-In CHO). Cells were monitored by microscopy and the size distribution of cell aggregates was quantified using a micro-flow imaging (MFI) apparatus. EC1-7GFP showed cell aggregation ability in the presence of $CaCl_2$. In contrast, EC1-7F224AGFP and mock cells did not show obvious cell aggregates in the presence of $CaCl_2$ (Fig. 4A, B). From this result, it was revealed that the homodimerization of LI-cadherin was necessary for LI-cadherin-dependent cell adhesion and that F224 was crucial for homodimerization also in the cellular environment.

Calcium-free linker

We next looked into the calcium-free linker between EC2 and EC3. Classical cadherins generally adopt a crescent-like shape (17, 18). However, in LI-cadherin, the arch-shape was disrupted at the calcium-free linker region and because of that EC1-4 exhibited unique alternating positioning of EC1-2 with respect to EC3-4. There have been several reports on cadherins possessing a calcium-free linker from which a biological role was suggested to be conferred by the flexibility of those linkers (19, 20).

To compare the rigidity of the canonical linker with three calcium ions and the calcium-free linker in LI-cadherin, we performed additional MD simulations. In addition to the monomeric states (Fig. S4), we also used the structure of the EC1-4 dimer (Fig. S7) as the initial structure of the simulations. We compared the rigidity of the linkers by calculating the RMSD values of Cα atoms of EC1 and EC3,

respectively, after superposing those of EC2 domain alone (Fig. 5A, B). Compared to the other domains, the EC3 domain in the monomer conformation exhibited the largest RMSD. EC3 in the homodimer also showed greater RMSD values than EC1, however, the values were significantly smaller than those of EC3 in the monomer form. Dihedral angles consisting of Cα atoms of residues at the edge of each domain also indicated that the EC1-4 monomer bends largely at the Ca^{2+}-free linker (Fig. S8A-C). These results showed that the calcium-free linker between EC2 and EC3 is more flexible than the canonical linker (Movie 3, 4).

However, even at the most pronounced bent position, the angle described was not as large as that in the calcium-free linker of *Drosophila* N-cadherin (DN-cadherin) (20). Our MD simulations of EC1-4 monomer suggested that the residue Asn237 located at the center of the calcium-free linker engaged in multiple H-bonds with peripheral residues (Fig. 5C). The stabilization energy of these H-bonds, involving residues in both EC2 and EC3, may explain why the LI-cadherin molecule cannot bend as sharply as DN-cadherin in the calcium-free region.

Another unique characteristic in the region surrounding the calcium-free linker was the existence of an α-helix at the bottom of EC2. To our best knowledge, this element at the bottom of the EC2 domain is not found in classical cadherins. The sequence alignment of the EC1-2 domains of human LI-, E-, N- and P-cadherin by ClustalW indicated that the insertion of the α-helix forming residues corresponded to the position immediately preceding the canonical calcium-binding motif DXE in classical cadherins (10) (Fig. S9). The Asp and Glu residues of the DXE motif in LI-cadherin dimer EC1 and EC3 coordinate with calcium ions (Fig. S10A, B) and was maintained throughout the simulation (Fig. S11). The α-helix in EC2 might compensate for the absence of calcium by conferring some rigidity to the molecule.

Solution structure of LI-cadherin EC1-2 and EC3-4

Our crystal structure of LI-cadherin from above has revealed an unprecedented homodimer architecture in the cadherin family. We analyzed the amino acid sequence homology of EC1-2 or EC3-4 of human LI-cadherin and EC1-2 of human classical cadherin (E-, N-, and P-cadherins) using ClustalW. We found that EC1-2 and EC3-4 of human LI-cadherin have 27~30% and 24~26% sequence homology with EC1-2 of human classical cadherin, respectively (Table S3). Because of the sequence homology with EC1-2 of classical cadherin, which preferentially forms X-dimers and strand swap dimers (15), we could not rule out the possibility that EC1-2 and EC3-4 of LI-cadherin might also adopt those conformations in solution. In particular, since EC3-4 has a Trp residue (Trp239) at its N-terminus, we hypothesized that when EC1-2 is absent, Trp239 may play the role of Trp2 in classical cadherins, i.e.

the key residue conferring strand swapping capabilities. Although EC1-2 does not possess the N-terminal Trp residue, we hypothesized that LI-cadherin EC1-2 could adopt an X-dimer as such of the classical T-cadherin. T-cadherin forms a durable X-dimer responsible for this cadherin's-dependent cell adhesion despite not displaying a Trp residue in the N-terminal position (21).

To evaluate whether EC1-2 and EC3-4 adopt the homodimer conformation or not, AUC was carried out. The K_D of dimerization of EC1-2 determined by AUC was 75 μM, whereas no dimerization was detected for EC3-4 (Fig. 6A, B). Interestingly, minor peaks consistent with a tetramer were observed in the measurement of EC1-2, although we could not determine the K_D value as the volume of tetramer included in the solution was very small. Small-angle X-ray scattering (SAXS) was further employed to analyze the homodimer structures in solution. Fig. 6C and 6D show the scattering curves of each construct. The values of the radius of gyration (Rg) calculated by AUTORG (22) were very similar to each other: 32.63 Å for EC1-2 and 32.97 Å for EC3-4, respectively (Fig. S12A, B). We separately extracted the monomeric forms of EC1-2 and EC3-4 from our structure of the EC1-4 dimer crystal to calculate the theoretical data and the dimerization experimental data.

To further examine the solution structures of the EC1-2 and EC3-4 regions at the molecular level, MD simulations of each construct were performed for more than 200 ns. Because the crystal structure lacked some of the N- and C-terminal residues and large portions of the sugar chains, MODELLER 9.18 (23) was employed to model the missing residues prior to the simulations, CHARMM-GUI Glycan Modeler (24) was then used to model the sugar chains (Fig. S12C), which were modelled as G0F. The simulated structures were extracted every 100 ps and clustered based on RMSD values of Cα atoms using GROMACS tool. Cutoff value was set to generate 10 to 15 clusters. The cluster including the maximum number of structures was selected and its center structure was defined as the theoretical structure of EC1-2 and EC3-4. We named the modified crystal structures EC1-2_plus and EC3-4_plus. Theoretical Rg values (Rg_{theo}) of EC1-2_plus and EC3-4_plus calculated in the FoXS server (25, 26) were 28.95 Å and 31.88 Å, respectively. At the same time, and consistent with the results of AUC, the experimental Rg value (Rg_{exp}) of EC1-2 was larger than that of Rg_{theo} calculated from EC1-2_plus. With respect to EC3-4, the value of Rg_{exp} was close to that of Rg_{theo} of EC3-4_plus. Moreover, the FoXS server was used to analyze the agreement of the scattering curves and the theoretical curves. The theoretical curve of the monomer of EC1-2_plus did not superpose well with the experimental scattering curve of EC1-2 (χ^2 = 1065.89) (Fig. 6C). In contrast, the theoretical curve of EC3-4_plus exhibited significant agreement with the experimental scattering curve of EC3-4 (χ^2 = 2.99) (Fig. 6D).

To investigate if LI-cadherin EC1-2 forms X-dimer or strand-swap dimer in solution, we generated the corresponding models of LI-cadherin in the X-dimer and strand-swap conformations by aligning EC1-2_plus to the crystal structures of E-cadherin X-dimer (PDB ID: 1EDH, (11)) and strand-swap dimer (PDB ID: 2O72, (27)). The generated structures were named EC1-2_plus_X and EC1-2_plus_ss, respectively (Fig. 7A). The scattering curve of EC1-2 did not fit well with any of the theoretical curves generated for EC1-2_plus, EC1-2_plus_X or EC1-2_plus_ss (Fig. 6C). However, the χ^2 value was smallest when the scattering curve was fit with the X-dimer, suggesting some similarities between the dimer of EC1-2 of LI-cadherin and the X-dimer of E-cadherin.

To elucidate the shape of the EC1-2 dimer, we performed *ab initio* bead modelling using DAMMIF (Fig. 7B, Fig. S12D) (28). The bead model suggested that two monomer molecules were interacting in a parallel or in an antiparallel fashion (Fig. 7C). Considering that LI-cadherin monomer bends at the calcium-free linker, as described above in the MD simulations, there is a possibility that the EC1-2 dimer observed by SAXS corresponds to another type of dimer. Whether both types of dimers are formed in human cells or not still remains elusive.

Homodimerization of LI-cadherin on cells

We next performed cell aggregation assays using CHO cells expressing various constructs of LI-cadherin in which domains were deleted, to elucidate the mechanism of cell-adhesion induced by LI-cadherin. LI-cadherin EC1-5 and EC3-7 expressing cells were separately established (EC1-5GFP and EC3-7GFP) (Table S1, Fig. S6). Importantly, neither EC1-5 nor EC3-7 expressing cells showed cell aggregation ability in the presence of $CaCl_2$ (Fig. 8).

The fact that EC1-5 expressing cells did not aggregate suggests that the dimer of EC1-5 detected by AUC was not efficiently replicated in a cell environment. Several reasons may explain that observation, such as steric hindrance between EC1 and the cell membrane (Fig. S13A), or inappropriate orientation of the protein molecules (Fig. S13B). It also indicated that the EC1-2 dimer observed by SAXS could not be formed because of instability of the molecule resulting from the high flexibility of the Ca^{2+}-free linker (Fig. S13C, D).

Expression of EC3-7 on the surface of the cells did not result in cell aggregation, an observation agreeing with the results of AUC and SAXS, which shows that EC3-4 does not form a dimer. The truncation of EC1-2 from LI-cadherin generates cadherin similar to classical cadherin in the point of view that it has five extracellular domains and that it has a Trp residue at the N-terminus. Together with the crystal structure of EC1-4 dimer, which showed that Trp239 was buried in its own hydrophobic pocket and was not participating in

homodimerization (Fig. S14), the fact that LI-cadherin EC3-7 did not aggregate represents a unique dimerization mechanism in LI-cadherin.

EC1-5 and EC3-7 expressing cells did not show aggregation ability even when they were mixed in equal amounts (Fig. S15). This result excluded the possibility of nonsymmetrical interaction of the domains (e.g. EC1-2 and E3-4, EC1-2 and EC6-7, etc.).

Discussion

Our study has revealed the architecture of the LI-cadherin homodimer, which differs from that of other cadherins, and the existence of a calcium-free linker between EC2 and EC3. We also showed that (i) LI-cadherin EC1-2 homodimer structure is different from that of the X-dimer and (ii) that LI-cadherin EC3-4 does not form homodimer. Our crystal structure also indicated that two additional domains have conferred unique features to LI-cadherin to play roles which cannot be achieved by the dimerization mechanism of classical cadherins. Based on the collective data gathered in this study, several trans-dimerization models of LI-cadherin on cells were elaborated (Fig. 9A-C).

We still could not determine whether EC1-2 homodimer observed by SAXS and AUC is the intermediate of the homodimer structure leading to the dimer observed in the crystal structure, or two types of homodimers are formed distinctively. Also, it is still unclear if EC1-2 molecules are aligned in a parallel or in an antiparallel arrangement when they homodimerize since the resolution of the bead model generated from SAXS data was insufficient to unambiguously assign EC1 from EC2. To date, there is no report about cadherins dimerizing in the same architecture as that of LI-cadherin. However, we predict that Ksp-cadherin, another member of the 7D-cadherin family (29) shares some of the characteristics of LI-cadherin, such as displaying seven extracellular domains and lacking calcium binding motif between domains EC2 and EC3 (Fig. S16). Since Ksp-cadherin also lacks calcium-binding motifs between EC5 and EC6, and between EC6 and EC7, we may expect Ksp-cadherin being more flexible than LI-cadherin.

If each type of homodimer (EC1-4 and EC1-2) was formed separately, that would suggest each homodimer should have a distinctive function. We speculate that intercellular distance and/or extracellular environment may influence the homodimerization state of LI-cadherin. For example, previous research has shown that LI-cadherin exists at the intercellular cleft of intestinal cells (7) and that trans-dimerization of LI-cadherin is required for water transport (8). The study indicated that the intercellular distance widens when water transport is inhibited. Although the intercellular distance in human cells is unknown, there is a possibility that the EC1-4 dimer (Fig. 9A) is formed when a high level of water transport is necessary (short intercellular distance) and EC1-2

dimer (Fig. 9B) appears when low level or no water transport is necessary (long intercellular distance). We also considered another possibility: High-affinity EC1-4 dimer (Fig. 9A) would be favored in normal intestinal cells to achieve water transport properly, whereas the low-affinity EC1-2 dimer (Fig. 9B, C) would be more abundant on metastatic cancer cells leading to their adhesion and to greater migration of cancer cells by weak cell adhesion.

Another notable difference between LI-cadherin and classical cadherins was the role of the cytoplasmic domain and the establishment of cis-interactions. Interaction of the cytoplasmic domain of classical cadherins with catenin is necessary for efficient cell-adhesion of classical cadherins (30, 31). Binding of classical cadherin cytoplasmic domain to F-actin mediated by catenin is essential for the clustering of cadherin molecules on cells leading to cis-interaction of cadherin molecules (32). A previous study has shown that the disruption of cis-interface in E-cadherin abolishes the ordered layer of E-cadherin molecules on liposomes, suggesting that the cis-interface is necessary for the formation of adhesive junction (18). In contrast, LI-cadherin cytoplasmic domain does not require any interaction with other proteins to exhibit cell adhesion ability (12). Moreover, cis-interaction of classical cadherins were also observed in the crystal lattices of C-, E-, and N-cadherin (17, 18). In contrast, we did not observe contacts which would suggest a cis-interface in our crystal structure. In addition to the lack of the conserved cytoplasmic domain, we believe that the existence of at least four N-glycans per monomer which extend to the opposite side of the dimer interface might inhibit the contact of other molecules. These features, together with additional two domains at N-terminus have conferred a unique cell adhesion mechanism by LI-cadherin.

In summary, our study has given new insights into the dimerization mechanism of LI-cadherin based on the crystal structure of LI-cadherin EC1-4, and simultaneously we conducted the characterization of LI-cadherin at the molecular level. We expect that our findings will accelerate the development of therapeutic agents targeting LI-cadherin, and help to understand the role of LI-cadherin in the progression of metastasis in certain cancers.

Experimental procedures

Protein sequence

Amino acid sequence of recombinant protein and LI-cadherin expressing CHO cells are summarized in Table S1.

Expression and purification of recombinant LI-cadherin

All LI-cadherin constructs were expressed and purified using the same method. All constructs were cloned in pcDNA 3.4 vector (ThermoFisher Scientific). Recombinant protein was expressed

using Expi293F™ Cells (ThermoFisher Scientific) following manufacturer's protocol. Cells were cultured for three days after transfection at 37 °C and 8% CO_2. The supernatant was collected and filtered followed by dialysis against a solution composed of 20 mM Tris-HCl at pH 8.0, 300 mM NaCl, and 3 mM $CaCl_2$. Immobilized metal affinity chromatography was performed using Ni-NTA Agarose (Qiagen). Protein was eluted by 20 mM Tris-HCl at pH 8.0, 300 mM NaCl, 3 mM $CaCl_2$, and 300 mM Imidazole. Final purification was performed by size exclusion chromatography (SEC) using HiLoad 26/600 Superdex 200 pg column (GE Healthcare) at 4 °C equilibrated in buffer A (10 mM HEPES-NaOH at pH 7.5, 150 mM NaCl, and 3 mM $CaCl_2$). Unless otherwise specified, samples were dialyzed against buffer A before analysis and filtered dialysis buffer was used for assays.

Crystallization of LI-cadherin EC1-4

Purified LI-cadherin EC1-4 was dialyzed against 10 mM HEPES-NaOH at pH 7.5, 30 mM NaCl, and 3 mM $CaCl_2$. After the dialysis, the protein was concentrated to 314 µM. Optimal condition for crystallization was screened using an Oryx8 instrument (Douglas Instruments) using commercial screening kits (Hampton Research). The crystal used for data collection was obtained in a crystallization solution containing 200 mM sodium sulfate decahydrate and 20% w/v Polyethylene glycol 3,350 at 20 °C. Suitable crystals were harvested, briefly incubated in mother liquor supplemented with 20% glycerol, and transferred to liquid nitrogen for storage until data collection.

Data collection and refinement

Diffraction data from a single crystal EC1-4 were collected in beamline BL-5A at the Photon Factory (Tsukuba, Japan) under cryogenic conditions (100 K). Diffraction images were processed with the program MOSFLM and merged and scaled with the program SCALA (33) of the CCP4 suite (34). The structure of the WT protein was determined by the molecular replacement method using the coordinates of P-cadherin (PDB entry code 4ZMY) (35) with the program PHASER (36). The models were refined with the programs REFMAC5 (37) and built manually with COOT (38). Validation was carried out with PROCHECK (39). Data collection and structure refinement statistics are given in Table 1. UCSF Chimera was used to render all of the molecular graphics (40).

Sedimentation velocity analytical ultracentrifugation (SV-AUC)

SV-AUC experiments were conducted using the Optima AUC (Beckman Coulter) equipped with an 8-hole An50 Ti rotor at 20 °C with 1, 2.5, 5, 10, 20, 40, and 60 µM of EC1-2, EC3-4, EC1-4 and EC1-5, dissolved in buffer A. Protein sample (390 µL) was loaded into the sample sector of a cell equipped

with sapphire windows and 12 mm double-sector charcoal-filled upon centerpiece. A volume of 400 µL of buffer was loaded into the reference sector of each cell. Data were collected at 42,000 rpm with a radial increment of 10 µm using a UV detection system.

The collected data were analyzed using continuous $c(s)$ distribution model implemented in program SEDFIT (version 16.2b) (41) fitting for the frictional ratio, meniscus, time-invariant noise, and radial-invariant noise using a regularization level of 0.68. The sedimentation coefficient ranges of 0-15 S were evaluated with a resolution of 150. The partial specific volumes of EC1-2, EC3-4, EC1-4 and EC1-5 were calculated based on the amino acid composition of each sample using program SEDNTERP 1.09 (42) and were 0.730 cm^3/g, 0.733 cm^3/g, 0.732 cm^3/g, and 0.734 cm^3/g, respectively. The buffer density and viscosity were calculated using program SEDNTERP 1.09 as 1.0055 g/cm^3 and 1.025 cP, respectively. Figures of $c(s_{20,w})$ distribution were generated using program GUSSI (version 1.3.2) (43). The weight-average sedimentation coefficient of each sample was calculated by integrating the range of sedimentation coefficients where peaks with obvious concentration dependence were observed. For the determination of the dissociation constant of monomer-dimer equilibrium, K_D, the concentration dependence of the weight-average sedimentation coefficient was fitted to the monomer-dimer self-association model implemented in program SEDPHAT (version 15.2b) (44).

Site-directed mutagenesis

Introduction of mutation to plasmid was performed as described previously (45).

Size exclusion chromatography with multi-angle light scattering (SEC-MALS)

The molecular weight of LI-cadherin was determined using superose 12 10/300 GL column (GE Healthcare) or Superose 6 Increase 10/300 GL column (GE Healthcare) with inline DAWN8+ multi angle light scattering (MALS) (Wyatt Technology), UV detector (Shimadzu), and refractive index (RI) detector (Shodex). Protien samples (45 µL) were injected at 100 µM or 50 µM. Analysis was performed using ASTRA software (Wyatt Technology). The protein conjugate method was employed for the analysis as sugar chains were bound to LI-cadherin. All detectors were BSA (Sigma-Aldrich).

MD simulation

Molecular dynamics simulations of LI-cadherin were performed using GROMACS 2016.3 (46) with the CHARMM36m force field (47). A whole crystal structure of EC1-4 homodimer, EC1-4 monomer form, EC1-2 monomer form and EC3-4 monomer form was used as the initial structure of the simulations, respectively. Missing residues

were modelled by MODELLER 9.18 (48). Solvation of the structures were performed with TIP3P water (49) in a rectangular box such that the minimum distance to the edge of the box was 15 Å under periodic boundary conditions through the CHARMM-GUI (50). Addition of N-bound type sugar chains (G0F) and the mutation of Phe224 in EC1-4 monomer to Ala224 were also performed through the CHARMM-GUI (24, 50). The protein charge was neutralized with added Na or Cl, and additional ions were added to imitate a salt solution of concentration 0.15 M. Each system was energy-minimized for 5000 steps and equilibrated with the NVT ensemble (298 K) for 1 ns. Further simulations were performed with the NPT ensemble at 298 K. The time step was set to 2 fs throughout the simulations. A cutoff distance of 12 Å was used for Coulomb and van der Waals interactions. Long-range electrostatic interactions were evaluated by means of the particle mesh Ewald method (51). Covalent bonds involving hydrogen atoms were constrained by the LINCS algorithm (52). A snapshot was saved every 10 ps. All trajectories were analyzed using GROMACS tools. RMSD, the number of hydrogen bonds, dihedral angles, distances between two atoms and clustering were computed by rms, hbond, gangle, distance and cluster modules, respectively.

The convergence of the trajectories was confirmed by calculating RMSD values of Cα atoms (Fig. S4A, B and S7A, B). As the molecule showed high flexibility at Ca^{2+}-free linker, as for EC1-4 WT monomer, EC1-4 F224A monomer and EC1-4 dimer, RMSD of each domain was calculated individually. Five Cα atoms at N-terminus was excluded from the calculation of RMSD of EC1 and EC1-2 as they were disordered. As the RMSD values were stable after running 20 ns of simulations, we did not consider the first 20 ns when we analyzed the trajectories.

Comparison of thermal stability by DSC

DSC measurement was performed using MicroCal VP-Capillary DSC (Malvern). The measurement was performed from 10 °C to 100 °C at the scan rate of 1 °C min^{-1}. Data was analyzed using Origin7 software.

Establishment of CHO cells expressing LI-cadherin

The DNA sequence of monomeric GFP was fused at the C-terminal of all human LI-cadherin constructs of which stable cell lines were established and was cloned in pcDNA™5/FRT vector (ThermoFisher Scientific). CHO cells stably expressing LI-cadherin-GFP were established using Flp-In™-CHO Cell Line following the manufacturer's protocol (ThermoFisher Scientific). Cloning was performed by the limiting dilution-culture method. Cells expressing GFP were selected and cultivated. Observation of the cells were performed by In Cell Analyzer 2000 (GE Healthcare). The cells were cultivated in Ham's F-

12 Nutrient Mixture (ThermoFisher Scientific) supplemented with 10 % fetal bovine serum (FBS), 1% L-Glutamine or 1% GlutaMAX™-I (ThermoFisher Scientific), 1% penicillin-streptomycin, and 0.5 mg mL^{-1} Hygromycin B at 37 °C and 5.0% CO_2.

Cell Imaging

Cells (100 µL) were added to a 96-well plate (Greiner) at 1 x 10^5 cells mL^{-1} and cultured overnight. After washing the cells with wash medium (Ham's F-12 Nutrient Mixture (ThermoFisher Scientific) supplemented with 10 % fetal bovine serum (FBS), 1% GlutaMAX™-I, 1% penicillin-streptomycin), Hoechst 33342 (ThermoFisher Scientific) (100 µL) was added to each well at 0.25 µg mL^{-1}. The plate was incubated at room temperature for 30 minutes. Cells were washed with wash medium twice and with 1x HMF (10 mM HEPES-NaOH at pH 7.5, 137 mM NaCl, 5.4 mM KCl, 0.34 mM Na_2HPO_4, 1 mM $CaCl_2$, and 5.5 mM glucose) twice. After that, 1x HMF (200 µL) was loaded to each well and the images were taken with an In Cell Analyzer 2000 instrument (GE Healthcare) using the FITC filter (490/20 excitation, 525/36 emission) and the DAPI filter (350/50 excitation, 455/50 emission) with 60 x 0.70 NA objective lens (Nikon).

Cell aggregation assay

Cell aggregation assay was performed by modifying the methods described previously (53, 54). Cells were detached from cell culture plate by adding 1x HMF supplemented with 0.01% trypsin and placing on a shaker at 80 rpm for 15 minutes at 37 °C. FBS was added to the final concentration of 20% to stop the trypsinization. Cells were washed with 1x HMF supplemented with 20% FBS once and with 1x HMF twice to remove trypsin. Cells were suspended in 1x HMF at 1 x 10^5 cells mL^{-1}. 500 µL of the cell suspension was loaded into 24-well plate coated with 1% w/v BSA. EDTA was added if necessary. After incubating the plate at room temperature for 5 minutes, 24-well plate was placed on a shaker at 80 rpm for 60 minutes at 37 °C. Images of the cells were taken using EVOS® XL Core Imaging System (ThermoFisher Scientific).

Micro-Flow Imaging (MFI)

Micro-Flow Imaging (Brightwell Technologies) was used to count the particle number and to visualize the cell aggregates after cell aggregation assay. After the cell aggregation assay described above, the plate was incubated at room temperature for 10 minutes and 500 µL of 4% Paraformaldehyde Phosphate Buffer Solution (Nacalai Tesque) was loaded to each well. Cells were injected to MFI after incubating the plate on ice for more than 20 minutes.

Solution structure analysis using SAXS

All measurements were performed at beamline BL-10C (55) of the Photon Factory (Tsukuba, Japan).

The experimental procedure is described previously (56). Concentrations of EC1-2 and EC3-4 were 255 μM and 157 μM, respectively. Data were collected using a PILATUS3 2M (Dectris). For the measurement of EC1-2, a wavelength was 1.500 Å with a camera distance 103 cm. For the measurement of EC3-4, a wavelength was 1.488 Å with a camera distance 101 cm. As for EC1-2, exposure time was 120 seconds and raw data between s values of 0.0084 and 0.82 $Å^{-1}$ were measured ($s = 4\pi\sin\theta/\lambda$). As for EC3-4, exposure time was 60 seconds and raw data between s values of 0.010 and 0.84 $Å^{-1}$ were measured. The background scattering intensity of buffer was subtracted from each measurement. The scattering intensities of nine measurements were averaged to produce the scattering curve of EC1-2 and four of them were averaged to produce the scattering curve of EC3-4. Data are placed on an absolute intensity scale. Conversion factor was calculated based on the scattering intensity of water. Bead modelling of EC1-2 was performed using DAMMIF (28). Symmetry was set to P2. The particle distance distribution function was calculated using GNOM considering $s < 0.245$ (57). D_{max} was calculated to be 107 Å. As the shape of the molecule was classified as "flat" by DATCLASS (58), expected shape was set as flat. Fifteen bead models generated using DAMMIF were averaged and filtered using DAMAVER package (59). A normalized discrepancy (NSD) was 0.561 ± 0.026 Å. The calculation of the theoretical curves of SAXS and χ^2 values were performed using FoXS server (25, 26).

Data availability: The coordinates and structure factors of LI-Cadherin EC1-4 have been deposited in the Protein Data Bank with entry code 7CYM. All remaining data are contained within the article.

Acknowledgements

We thank Dr. S. Kudo and Dr. H. Akiba for expert advice. We thank Dr. O. Kusano-Arai, Dr. H. Iwanari and Dr. T. Hamakubo for providing us with gene sequence of LI-cadherin.

Funding and additional information

The supercomputing resources in this study were provided by the Human Genome Center at the Institute of Medical Science, The University of Tokyo, Japan. This work was funded by a Grant-in-Aid for Scientific Research (A) 16H02420 (K.T.), a Grant-in-Aid for Scientific Research (B) 20H02531 (K.T.) a Grant-in-Aid for Scientific Research on Innovative Areas 19H05760 and 19H05766 (K.T.) and a Grant-in-Aid for JSPS fellows (A.Y.). We are grateful to the staff of the Photon Factory (Tsukuba, Japan) for excellent technical support. Access to beamlines BL-5A and BL-10C was granted by the Photon Factory Advisory Committee (Proposal Numbers 2018G116 and 2017G661).

Conflict of Interest

The authors declare that they have no conflicts of interest with the contents of this article.

References

1. Takeichi Masatoshi (1988) The cadherins: Cell-cell adhesion molecules controlling animal morphogenesis. *Development*. **102**, 639–655
2. Van Roy, F. (2014) Beyond E-cadherin: Roles of other cadherin superfamily members in cancer. *Nat. Rev. Cancer*. **14**, 121–134
3. Hulpiau, P., and van Roy, F. (2009) Molecular evolution of the cadherin superfamily. *Int. J. Biochem. Cell Biol.* **41**, 349–369
4. Hinoi, T., Lucas, P. C., Kuick, R., Hanash, S., Cho, K. R., and Fearon, E. R. (2002) CDX2 regulates liver intestine-cadherin expression in normal and malignant colon epithelium and intestinal metaplasia. *Gastroenterology*. **123**, 1565–1577
5. Ko, S., Chu, K. M., Luk, J. M., Wong, B. W., Yuen, S. T., Leung, S. Y., and Wong, J. (2004) Overexpression of LI-cadherin in gastric cancer is associated with lymph node metastasis. *Biochem. Biophys. Res. Commun.* **319**, 562–568

6. Matsusaka, K., Ushiku, T., Urabe, M., Fukuyo, M., Abe, H., Ishikawa, S., Seto, Y., Aburatani, H., Hamakubo, T., Kaneda, A., and Fukayama, M. (2016) Coupling CDH17 and CLDN18 markers for comprehensive membrane-targeted detection of human gastric cancer. *Oncotarget*. **7**, 64168–64181
7. Berndorff, D., Gessner, R., Kreft, B., Schnoy, N., Lajous-Petter, A. M., Loch, N., Reutter, W., Hortsch, M., and Tauber, R. (1994) Liver-intestine cadherin: Molecular cloning and characterization of a novel Ca^{2+}-dependent cell adhesion molecule expressed in liver and intestine. *J. Cell Biol.* **125**, 1353–1369
8. Weth, A., Dippl, C., Striedner, Y., Tiemann-Boege, I., Vereshchaga, Y., Golenhofen, N., Bartelt-Kirbach, B., and Baumgartner, W. (2017) Water transport through the intestinal epithelial barrier under different osmotic conditions is dependent on LI-cadherin trans-interaction. *Tissue Barriers*. 10.1080/21688370.2017.1285390
9. Jung, R., Wendeler, M. W., Danevad, M., Himmelbauer, H., and Geßner, R. (2004) Phylogenetic origin of LI-cadherin revealed by protein and gene structure analysis. *Cell. Mol. Life Sci.* **61**, 1157–1166
10. Shapiro, L., Fannon, A. M., Kwong, P. D., Thompson, A., Lehmann, M. S., Gerhard, G., Als-Nielsen, J., Als-Nielsen, J., Colman, D. R., and Hendrickson, W. A. (1995) Structural basis of cell-cell adhesion by cadherins. *Nature*. **374**, 327–337
11. Nagar, B., Overduin, M., Ikura, M., and Rini, J. M. (1996) Structural basis of calcium-induced E-cadherin rigidification and dimerization. *Nature*. **380**, 360–364
12. Kreft, B., Berndorff, D., Böttinger, A., Finnemann, S., Wedlich, D., Hortsch, M., Tauber, R., and Gener, R. (1997) LI-cadherin-mediated cell-cell adhesion does not require cytoplasmic interactions. *J. Cell Biol.* **136**, 1109–1121
13. Brasch, J., Harrison, O. J., Honig, B., and Shapiro, L. (2012) Thinking outside the cell: How cadherins drive adhesion. *Trends Cell Biol.* **22**, 299–310
14. Nicoludis, J. M., Vogt, B. E., Green, A. G., Schärfe, C. P. I., Marks, D. S., and Gaudet, R. (2016) Antiparallel protocadherin homodimers use distinct affinity-and specificity-mediating regions in cadherin repeats 1-4. *Elife*. **5**, 1–14
15. Harrison, O. J., Bahna, F., Katsamba, P. S., Jin, X., Brasch, J., Vendome, J., Ahlsen, G., Carroll, K. J., Price, S. R., Honig, B., and Shapiro, L. (2010) Two-step adhesive binding by classical cadherins. *Nat. Struct. Mol. Biol.* **17**, 348–357
16. Krissinel, E., and Henrick, K. (2007) Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **372**, 774–797

17. Boggon, T. J., Murray, J., Chappuis-Flament, S., Wong, E., Gumbiner, B. M., and Shapiro, L. (2002) C-cadherin ectodomain structure and implications for cell adhesion mechanisms. *Science.* **296**, 1308–1313

18. Harrison, O. J., Jin, X., Hong, S., Bahna, F., Ahlsen, G., Brasch, J., Wu, Y., Vendome, J., Felsovalyi, K., Hampton, C. M., Troyanovsky, R. B., Ben-Shaul, A., Frank, J., Troyanovsky, S. M., Shapiro, L., and Honig, B. (2011) The extracellular architecture of adherens junctions revealed by crystal structures of type i cadherins. *Structure*. **19**, 244–256

19. Araya-Secchi, R., Neel, B. L., and Sotomayor, M. (2016) An elastic element in the protocadherin-15 tip link of the inner ear. *Nat. Commun.* 10.1038/ncomms13458

20. Jin, X., Walker, M. A., Felsövályi, K., Vendome, J., Bahna, F., Mannepalli, S., Cosmanescu, F., Ahlsen, G., Honig, B., and Shapiro, L. (2012) Crystal structures of Drosophila N-cadherin ectodomain regions reveal a widely used class of Ca^{2+}-free interdomain linkers. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E127–E134

21. Ciatto, C., Bahna, F., Zampieri, N., Vansteenhouse, H. C., Katsamba, P. S., Ahlsen, G., Harrison, O. J., Brasch, J., Jin, X., Posy, S., Vendome, J., Ranscht, B., Jessell, T. M., Honig, B., and Shapiro, L. (2010) T-cadherin structures reveal a novel adhesive binding mechanism. *Nat. Struct. Mol. Biol.* **17**, 339–347

22. Petoukhov, M. V., Konarev, P. V., Kikhney, A. G., and Svergun, D. I. (2007) ATSAS 2.1 - Towards automated and web-supported small-angle scattering data analysis. *J. Appl. Crystallogr.* **40**, 223–228

23. Sali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815

24. Jo, S., Song, K. C., Desaire, H., MacKerell, A. D., and Im, W. (2011) Glycan reader: Automated sugar identification and simulation preparation for carbohydrates and glycoproteins. *J. Comput. Chem.* **32**, 3135–3141

25. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A., and Sali, A. (2013) Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys. J.* **105**, 962–974

26. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A., and Sali, A. (2016) FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res.* **44**, W424–W429

27. Parisini, E., Higgins, J. M. G., Liu, J. huan, Brenner, M. B., and Wang, J. huai (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**, 401–411

28. Franke, D., and Svergun, D. I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346
29. Wendeler, M. W., Praus, M., Jung, R., Hecking, M., Metzig, C., and Geßner, R. (2004) Ksp-cadherin is a functional cell-cell adhesion molecule related to LI-cadherin. *Exp. Cell Res.* **294**, 345–355
30. Nagafuchi, A., and Takeichi, M. (1988) Cell binding function of E-cadherin is regulated by the cytoplasmic domain. *EMBO J.* **7**, 3679–3684
31. Ozawa, M., Barbault, H., and Kemler, R. (1989) The cytoplasmic domain of the cell adhesion molecule uvomorulin assoicates with three independent proteins structurally related in different species. *EMBO J.* **8**, 1711–1717
32. Hong, S., Troyanovsky, R. B., and Troyanovsky, S. M. (2013) Binding to F-actin guides cadherin cluster assembly, stability, and movement. *J. Cell Biol.* **201**, 131–143
33. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 72–82
34. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 235–242
35. Kudo, S., Caaveiro, J. M. M., and Tsumoto, K. (2016) Adhesive Dimerization of Human P-Cadherin Catalyzed by a Chaperone-like Mechanism. *Structure*. **24**, 1523–1536
36. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674
37. Murshudov, G. N. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.* **53**, 240–255
38. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501
39. Laskowski, R. A. (1993) PROCHECK-a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291
40. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612
41. Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity

ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78**, 1606–1619

42. Laue, T. M., Shah, B., Ridgeway, T. M., and Pelletier, S. L. (1992) Computer-aided interpretation of analytical sedimentation data for proteins. *Anal. ultracentrifugation Biochem. Polym. Sci.*
43. Brautigam, C. A. (2015) Chapter Five - Calculations and Publication-Quality Illustrations for Analytical Ultracentrifugation Data. in *Methods in Enzymology* (Cole, J. L. ed), **562**, 109–133
44. Schuck, P. (2003) On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal. Biochem.* **320**, 104–124
45. Yui, A., Akiba, H., Kudo, S., Nakakido, M., Nagatoishi, S., and Tsumoto, K. (2017) Thermodynamic analyses of amino acid residues at the interface of an antibody B2212A and its antigen roundabout homolog 1. *J. Biochem.* 10.1093/jb/mvx050
46. Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. **1**–**2**, 19–25
47. Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., De Groot, B. L., Grubmüller, H., and MacKerell, A. D. (2016) CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nat. Methods*. **14**, 71–73
48. Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M., Pieper, U., and Sali, A. (2006) Comparative Protein Structure Modeling Using Modeller. *Curr. Protoc. Bioinforma.* **15**, 5.6.1-5.6.37
49. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935
50. Jo, S., Kim, T., Iyer, V. G., and Im, W. (2008) CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865
51. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092
52. Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997) LINCS: A Linear Constraint Solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472
53. Urushihara, H., Takeichi, M., Hakura, A., and Okada, T. S. (1976) Different cation requirements for aggregation of BHK cells and their transformed derivatives. *J. Cell Sci.* **22**, 685–695
54. Urushihara, H., Ozaki, H. S., and Takeichi, M. (1979) Immunological detection of cell surface components related with aggregation of Chinese hamster and chick embryonic cells. *Dev. Biol.* **70**, 206–216
55. Shimizu, N., Mori, T., Nagatani, Y., Ohta, H., Saijo, S., Takagi, H., Takahashi, M., Yatabe, K.,

Kosuge, T., and Igarashi, N. (2019) BL-10C, the small-angle x-ray scattering beamline at the photon factory. *AIP Conf. Proc.* 10.1063/1.5084672

56. Kudo, S., Caaveiro, J. M. M., Miyafusa, T., Goda, S., Ishii, K., Matsuura, T., Sudou, Y., Kodama, T., Hamakubo, T., and Tsumoto, K. (2012) Structural and thermodynamic characterization of the self-adhesive properties of human P-cadherin. *Mol. Biosyst.* **8**, 2050–2053
57. Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503
58. Franke, D., Petoukhov, M. V., Konarev, P. V., Panjkovich, A., Tuukkanen, A., Mertens, H. D. T., Kikhney, A. G., Hajizadeh, N. R., Franklin, J. M., Jeffries, C. M., and Svergun, D. I. (2017) ATSAS 2.8: A comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50**, 1212–1225
59. Volkov, V. V., and Svergun, D. I. (2003) Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36**, 860–864
60. Laskowski, R. A., and Swindells, M. B. (2011) LigPlot+: Multiple Ligand–Protein Interaction Diagrams for Drug Discovery. *J. Chem. Inf. Model.* **51**, 2778–2786

Abbreviations

LI-cadherin, Liver Intestine-cadherin; EC, extracellular cadherin; MD, molecular dynamics; SAXS, small angle X-ray scattering; AUC, analytical ultracentrifugation; SEC-MALS, size exclusion chromatography-multi angle light scattering; DSC, differential scanning calorimetry; MFI, micro-flow imaging; RMSD, root mean square deviation

Table 1: Data collection and refinement statistics.

Statistical values given in parenthesis refer to the highest resolution bin.

Data Collection	**LI-cadherin (EC1-4)**
Space Group	P 1 2_1 1
Unit cell	
a, b, c (Å)	80.36, 70.84, 134.22
α, β, γ (°)	90.0, 98.7, 90.0
Resolution (Å)	55.19 - 2.70 (2.85 - 2.70)
Wavelength	1.0000
Observations	252,491 (37,071)
Unique reflections	41,272 (5,955)
$R_{merge.}$	0.095 (0.858)
$R_{p.i.m.}$	0.041 (0.371)
$CC_{1/2}$	0.998 (0.907)
$I / \sigma (I)$	11.6 (1.8)
Multiplicity	6.1 (6.2)
Completeness (%)	99.9 (100.0)
Refinement Statistics	
Resolution (Å)	55.19 - 2.70
R_{work} / R_{free} (%)	22.2 / 27.4
No. protein chains	2
No. atoms	
Protein	6,834
Ca^{2+}	12
Water	33
B-factor (Å^2)	
Protein	78.2
Ca^{2+}	69.3
Water	62.2
Ramachandran Plot	
Preferred (%)	85.7
Allowed (%)	14.3
Outliers (%)	0
RMSD Bond (Å)	0.013
RMSD Angle (°)	1.83
PDB entry code	7CYM

Table 2: Results of SEC-MALS

Sample	SEC-MALS		DSC		
	Concentration (μM)	MW[1] (kDa)	Concentration (μM)	T_{M1} (°C)	T_{M2} (°C)
WT	100	53.7	7.8	60.2 ± 0.0[2]	N.D.[3]
	50	51.9			
I169A	100	50.5	7.9	61.7 ± 0.0	N.D.[3]
	50	50.1			
L171A	100	51.1	7.9	60.3 ± 0.0	N.D.[3]
	50	49.7			
N176A	100	53.1	7.5	57.7 ± 0.0	63.7 ± 0.0
	50	N.D.[3]			
V210A	100	51.5	7.9	56.5 ± 0.0	63.0 ± 0.0
	50	50.3			
N222A	100	52.2	7.4	58.7 ± 0.0	63.9 ± 0.0
	50	49.2			
F224A	100	49.5	8.1	54.2 ± 0.0	62.5 ± 0.0
	50	49.4			
L355A	100	54.8	7.8	59.9 ± 0.0	65.3 ± 0.1
	50	52.5			
N371A	100	53.1	7.1	59.5 ± 0.0	N.D.[3]
	50	50.7			
F376A	100	55.1	7.8	60.4 ± 0.0	N.D.[3]
	50	50.6			
Y399A	100	53.6	7.5	59.9 ± 0.0	N.D.[3]
	50	51.7			
Q404A	100	54.8	7.8	59.6 ± 0.0	N.D.[3]
	50	51.4			

1. The molecular weight of the protein does not include the glycan moiety. The theoretical molecular weight of EC1-4 wild type without glycan is 49.8 kDa.

2. T_m ± error is shown.

3. Not determined.

Figures

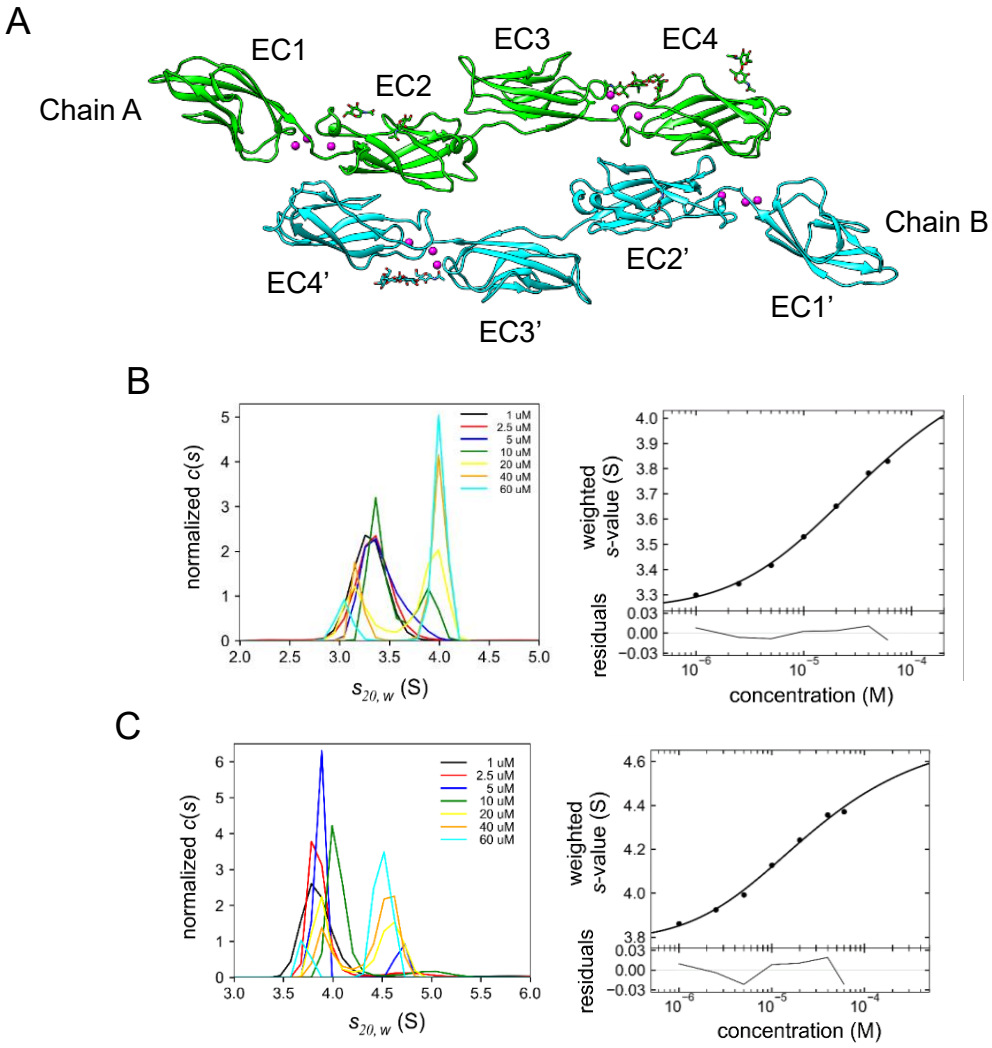

Figure 1. Structure and dimerization of human LI-cadherin. **A.** Crystal structure of EC1-4 showing an antiparallel homodimer architecture. Calcium ions are depicted in magenta. No calcium ions were observed between domains EC2 and EC3 in either molecule. Four partial N-glycans were modeled in chain A (light green) and two in chain B (cyan) (the amino acid sequence of EC1-4 is given in Table S1). **B, C.** Sedimentation plot of SV-AUC experiment of EC1-4 (B) and EC1-5 (C).

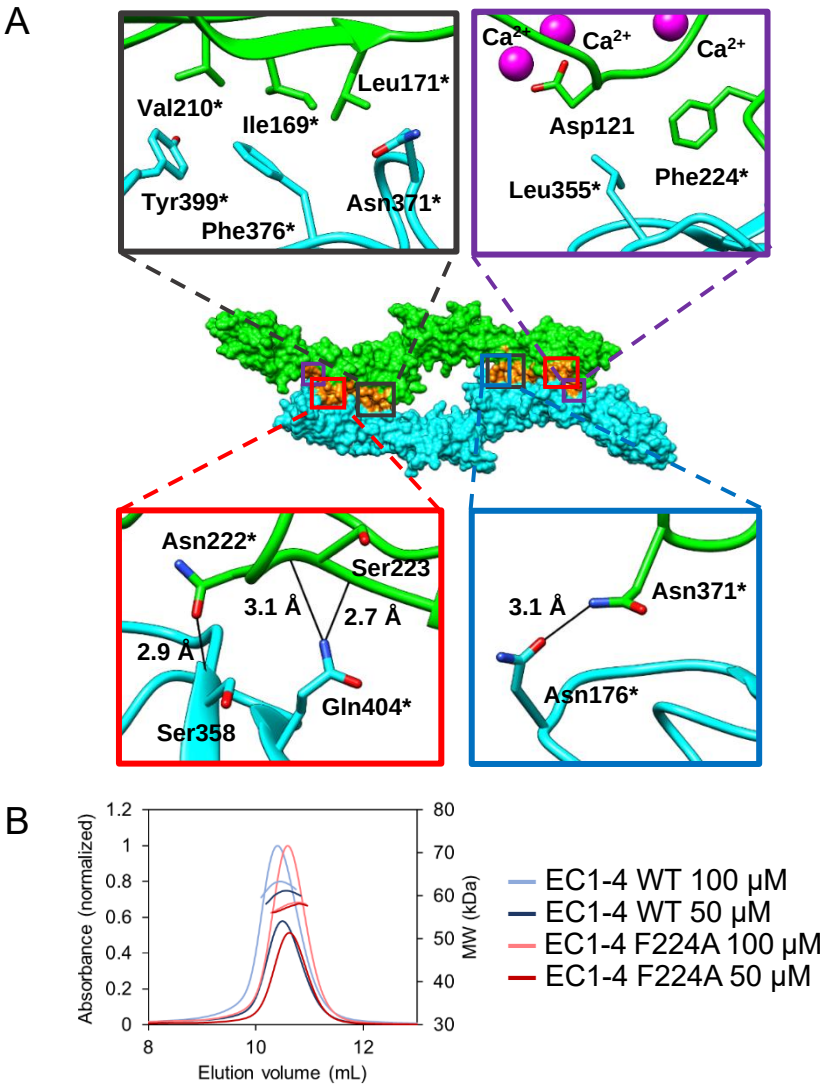

Figure 2. Homodimerization of LI-cadherin. **A.** Interface of the EC1-4 homodimer. The non-polar interaction residues are shown in black and purple rectangles (top panels). Residues involved in hydrogen bonds (black solid lines) are shown within the red and blue rectangles (bottom panel). Residues indicated with an asterisk were individually mutated to Ala to evaluate their contribution to dimerization. **B.** SEC-MALS chromatogram of EC1-4 WT and F224A. The volume at which the WT protein eluted and its MW increased in a concentration-dependent manner, whereas those values did not show concentration dependency in the mutant F224A. For each construct the absorbance of the larger peak of the sample at a concentration of 100 µM was normalized to a value of one.

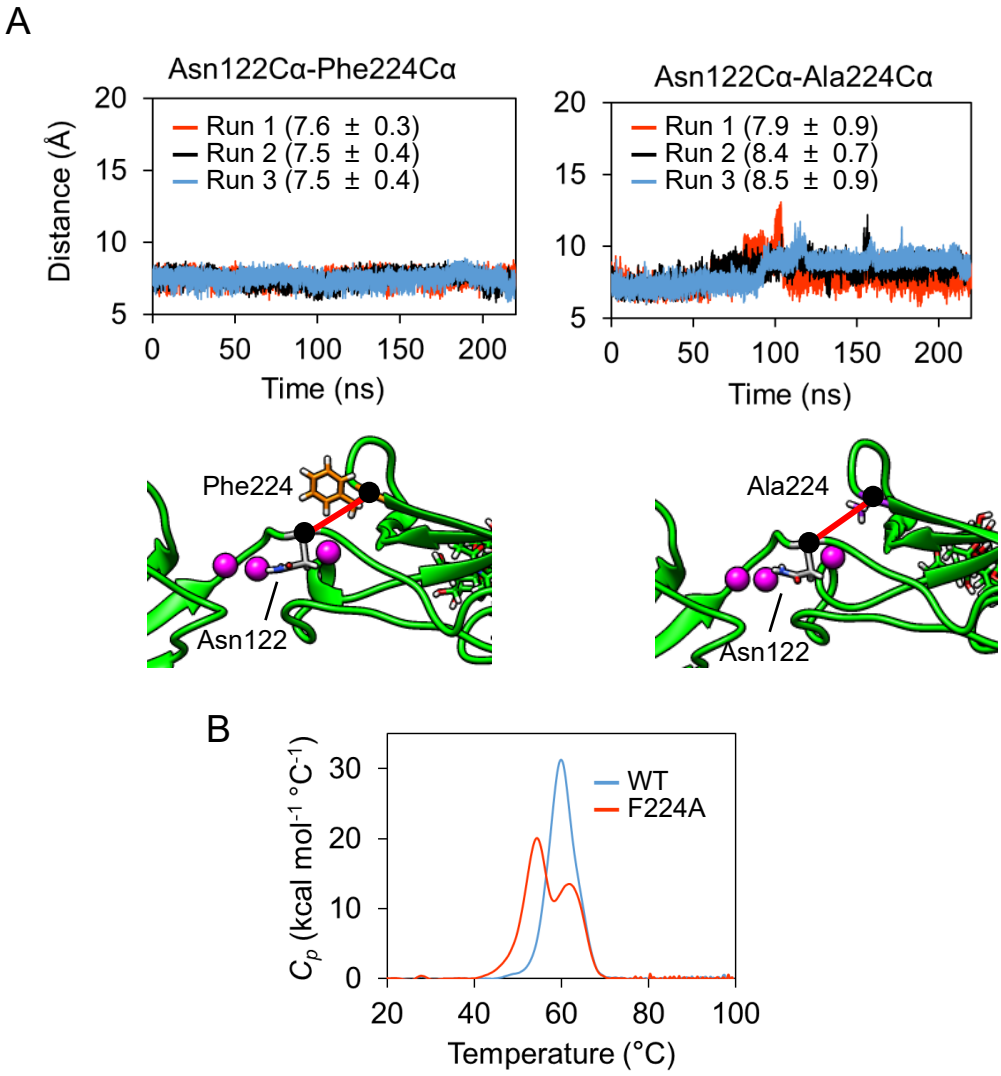

Figure 3. Mechanism of dimerization facilitated by Phe224. **A.** The distance between Phe224 (orange) Cα or Ala224 (purple) Cα and Asn122 (grey) Cα was evaluated by MD simulations. The Cα atoms are indicated by black circles. The distance calculated by the simulations is indicated with red line. Time course is shown on the portion of the panel at the upper part of each structure. Each MD simulations run is shown in red, black and blue. Averages and standard deviations from 20 to 220 ns of each simulation are shown in parentheses. **B.** Thermal stability of EC1-4 WT and F224A determined by differential scanning calorimetry. Two transitions appeared in the sample of F224A. The first transition at lower temperature seems to have appeared due to the loss of intramolecular interaction around the residue at position 224.

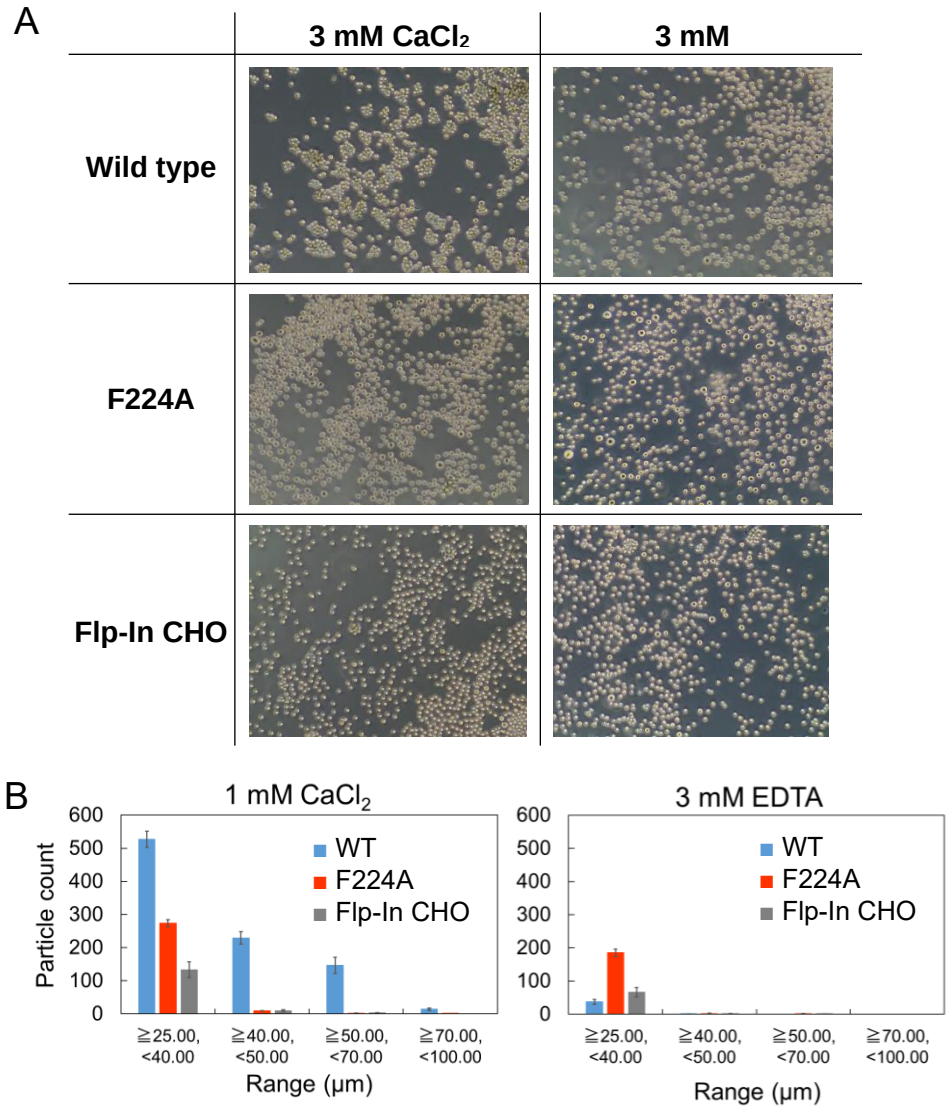

Figure 4. Cell aggregation assay. **A.** Cell adhesion was followed by the aggregation of CHO cells expressing LI-cadherin EC1-7 WT or EC1-7 F224A. Cell aggregation was studied in the presence of Ca^{2+} ions (left panels) or absence of Ca^{2+} (right panels). **B.** Size distribution of cell aggregates determined by MFI. Particles larger than 25 μm were regarded as cell aggregates. Only EC1-7 WT expressing cells in the existence of 1 mM $CaCl_2$ showed significant number of cell aggregates that were 40 μm or larger. Data shows the mean ± SEM of four measurements.

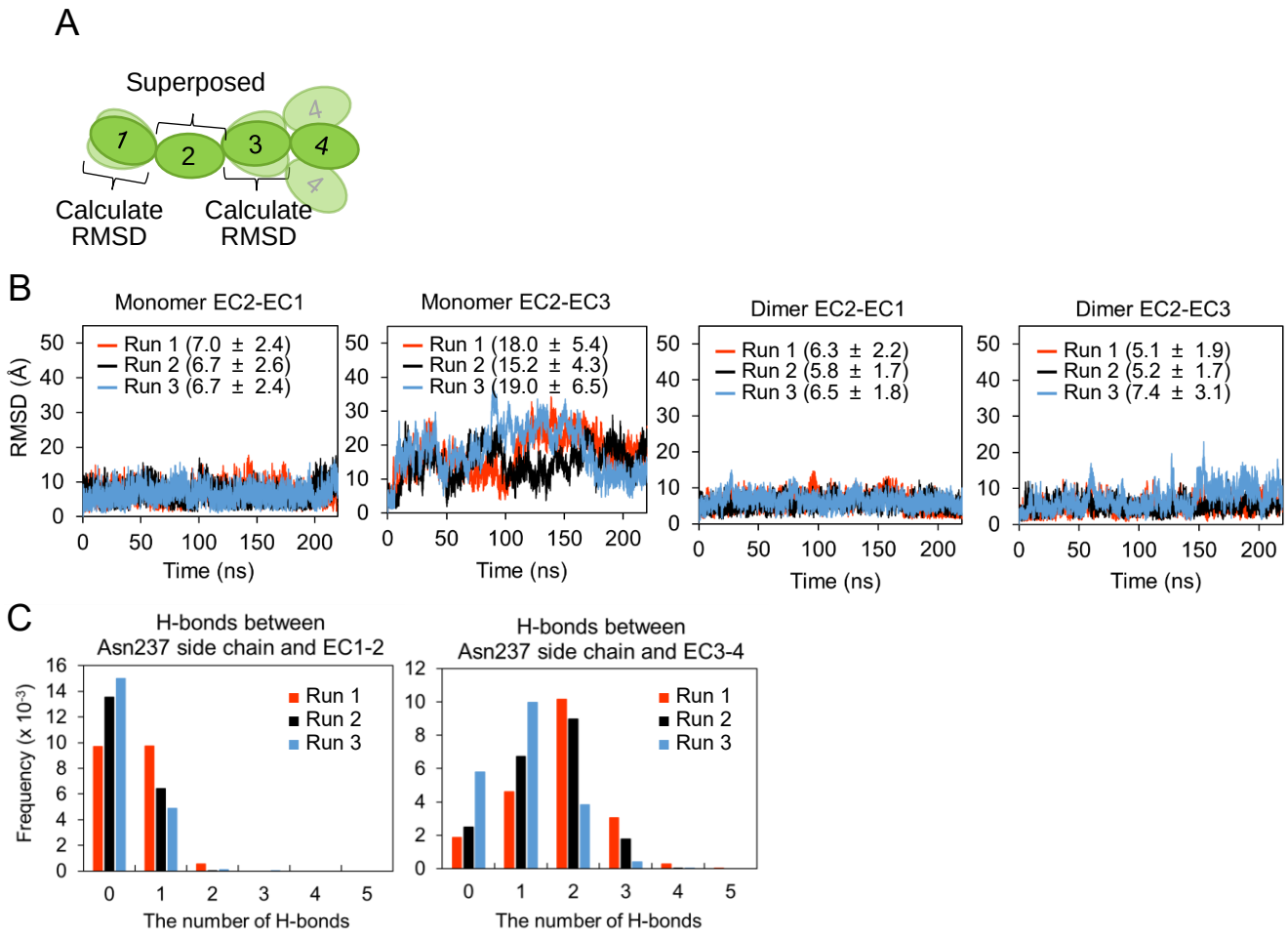

Figure 5. Calcium-free linker. **A.** Schematic view of how RMSD values were calculated. **B.** MD simulations of LI-cadherin. RMSD values of EC1 Cα or EC3 Cα against EC2 Cα. Chain A of EC1-4 dimer crystal structure was employed as the initial structure. **C.** RMSD values of EC1 Cα or EC3 Cα in chain A of the dimer against EC2 Cα in the chain A. **C.** Number of H-bonds between the side chain of Asn237 and other residues in EC1-2 or EC3-4.

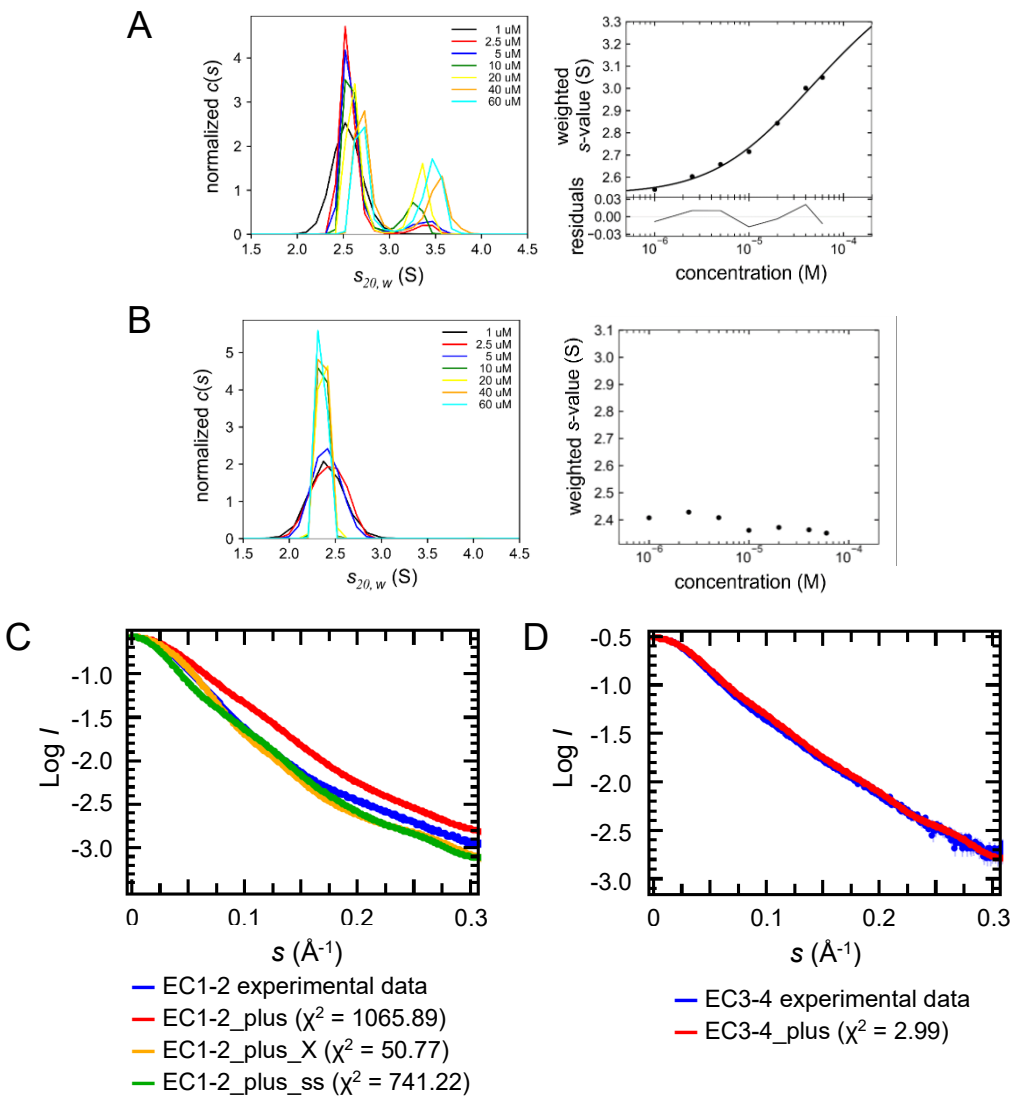

Figure 6. Dimerization of short constructs. **A, B.** Sedimentation plot of AUC experiment corresponding to EC1-2 (A) and EC3-4 (B). **C.** Scattering curve of LI-cadherin EC1-2 and theoretical curves of EC1-2_plus, EC1-2_plus_X and EC1-2_plus_ss. Calculation of theoretical curves and χ^2 values were performed using FoXS server ($s = 4\pi\sin\theta/\lambda$, I = Intensity). **D.** Scattering curve of LI-cadherin EC3-4 and theoretical curve of EC3-4_plus.

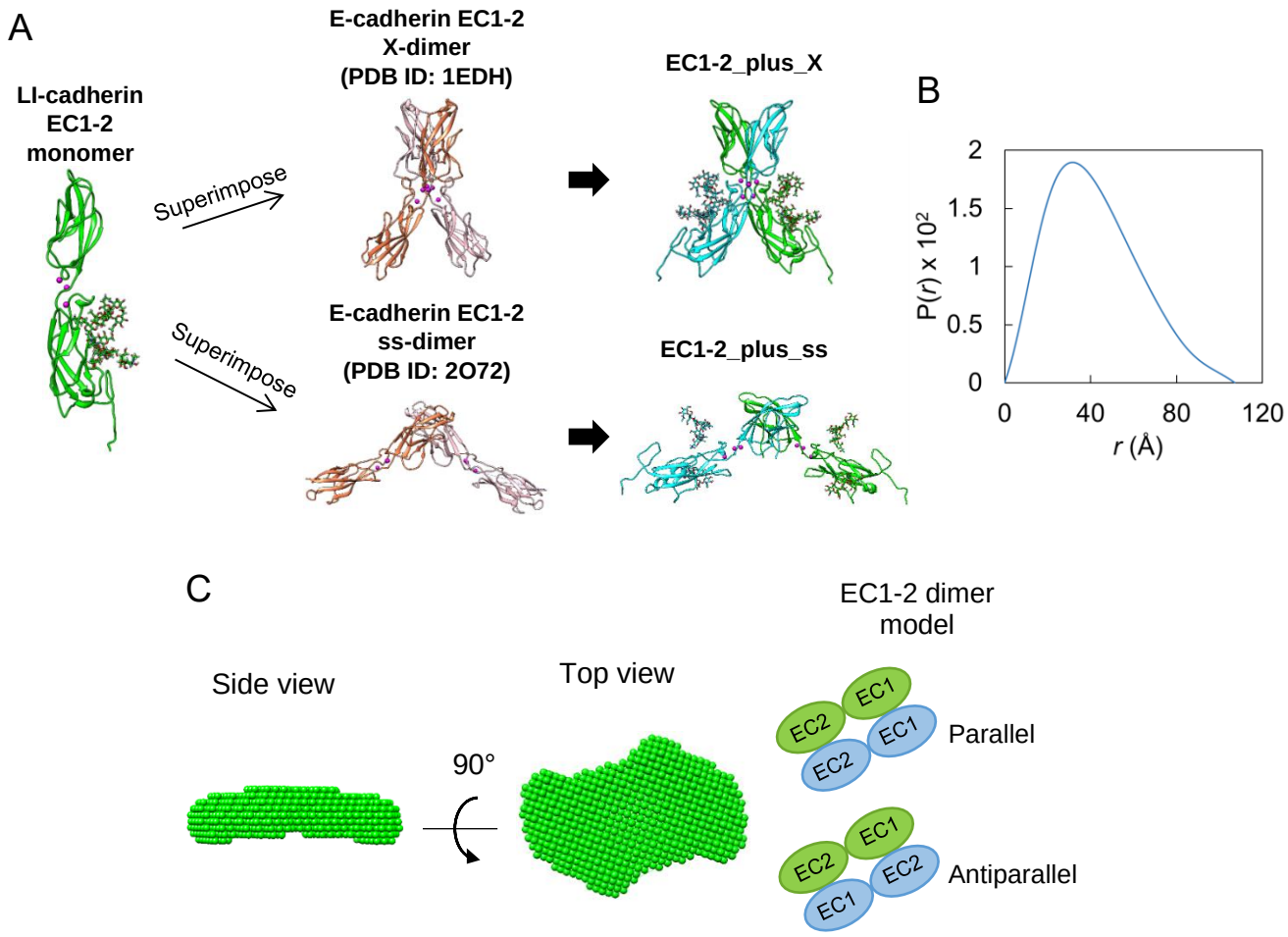

Figure 7. Modeling LI-cadherin dimer. **A.** Procedure to generate EC1-2_plus_X and _ss. **B.** Distance distribution function of EC1-2. **C.** *Ab initio* bead model of EC1-2 homodimer generated using DAMMIF. Side view and top view are shown. Proposed models of EC1-2 homodimer are shown on the right. Both parallel and antiparallel forms were suggested from the shape of the bead model.

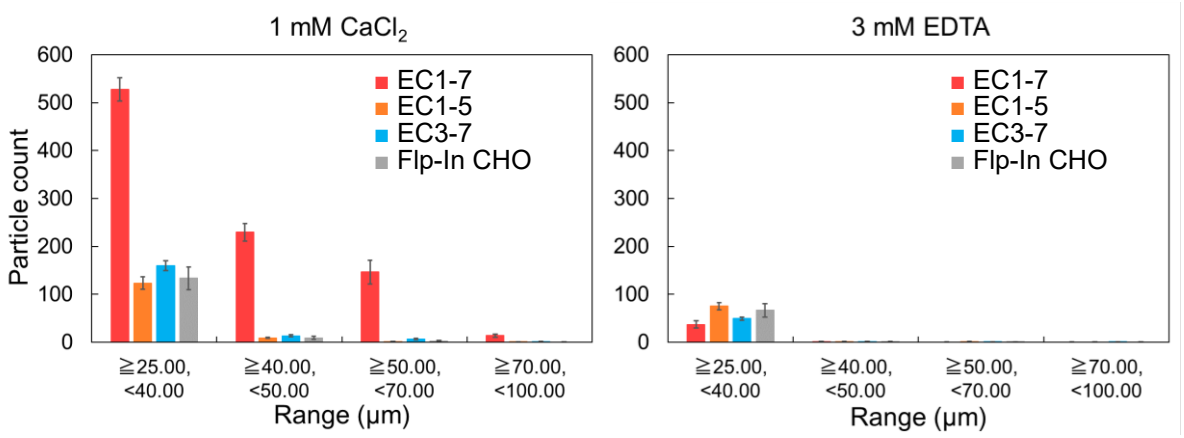

Figure 8. Cell adhesion mediated by short constructs. Cell aggregation assay using EC1-5GFP and EC3-7GFP. EC1-7GFP and Flp-In CHO (mock cells) were used as positive and negative control, respectively. Particles larger than 25 μm were considered as cell aggregates. The number of cell aggregates of both EC1-5GFP and EC3-7GFP in the presence or absence of Ca^{2+} were determined. Data show mean ± SEM of four measurements.

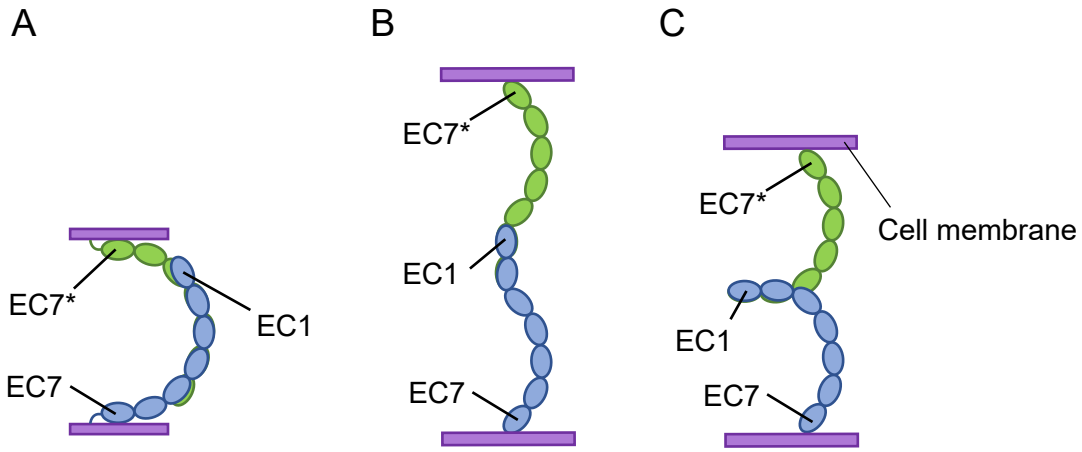

Figure 9. Models of interaction. **A-C.** EC1-7 homodimer models on cells. LI-cadherin extracellular domains are colored in blue and light green. Orientation of EC5-7 against EC1-4 was predicted by superposing EC1-2 of E-cadherin EC1-5 (PDB ID: 3Q2V, (18)) to LI-cadherin EC3-4.