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: Jose M. M. Caaveiro and Kouhei Tsumoto

E-mails: jose@phar.kyushu-u.ac.jp; 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 therapeutic molecules 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. Various biochemical techniques and molecular dynamics (MD) simulations were employed to elucidate the mechanism of homodimerization. We also showed that the formation of the homodimer observed by the crystal structure is necessary for LI-cadherin-dependent cell adhesion by performing cell aggregation assay.

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 composed of extracellular region which includes seven extracellular cadherin (EC) repeats, single transmembrane domain and a short cytoplasmic domain (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 of 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 cadherin repeats and the length and the sequence of the cytoplasmic domain. Classical cadherins possess five cadherin repeats 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 remain unknown. Homodimerization is the fundamental event in cadherin-mediated cell adhesion as has been shown previously (13, 14). For example, classical cadherins form a homodimer mediated by the interaction between their two N-terminal cadherin repeats (EC1-2) (10, 15).

In this study, we aimed to characterize LI-cadherin at the molecular level as the molecular characteristics of the target protein may be significant for the rational design of therapeutic approaches. We have extensively validated LI-cadherin to identify the homodimer architecture of LI-cadherin. Here, we report the crystal structure of human LI-cadherin EC1-4 homodimer. The 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, 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. Additionally, we showed that the EC1-4 homodimer is necessary for LI-cadherin-dependent cell adhesion through cell aggregation assays. 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

Investigation of the domains responsible for the homodimerization of LI-cadherin

In order to predict which extracellular cadherin (EC) repeats are responsible for the homodimerization of LI-cadherin, we compared the sequence of human LI-cadherin and human classical cadherins (E-, N- and P-cadherins) using ClustalW. As has been pointed out in the previous study (9), it was revealed that EC3-7 of human LI-cadherin has sequence homology with EC1-5 of human classical cadherins, and EC1-2 of human LI-cadherin has sequence homology with EC1-2 of classical cadherins **(Fig. 1)**. Notably, Trp239 locates at the N-terminus of LI-cadherin EC3 and it has been suggested that this Trp residue might function as the conserved Trp2 of classical cadherin EC1, which plays a crucial role in the formation of strand swap-dimer (ss-dimer) (9, 10, 16–18). Considering that EC1-2 of classical cadherins is responsible for homodimerization, we predicted that EC1-2 and EC3-4 of LI-cadherin, which have sequence homology with EC1-2 of classical

cadherins are responsible for its dimerization. Therefore, we first analyzed the homodimerization propensity of EC1-4, EC1-2, and EC3-4 **(Table S1)** of human LI-cadherin.

The dissociation constant (K_D) of the EC1-4 homodimer and the EC1-2 homodimer were determined by sedimentation velocity analytical ultracentrifugation (SV-AUC), obtaining values of 39.8 μM and 75.0 μM, respectively **(Fig. 2A)**. We did not observe dimer fraction when employing EC3-4 despite the sequence similarity with EC1-2 of classical cadherins and the presence of Trp239 in EC3 located at the analogous position to that of Trp2 in EC1 of classical cadherin **(Fig. 2A)**. The solution structure of EC3-4 was monomeric as determined by small angle X-ray scattering (SAXS), supporting the results of SV-AUC **(Fig. 2B, Fig. S1 and Table S2)**.

Crystal structure analysis of EC1-4 homodimer

We successfully obtained the X-ray crystal structure of EC1-4 at 2.7 Å resolution **(Fig. 3 and Table 1)**. Each 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. 3)**. We also observed four N-glycans and two N-glycans bound to chain A and B, respectively, as predicted from the 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 existence of a calcium-free linker between EC2 and EC3, and (ii) the architecture of the homodimer.

A previous study had suggested that LI-cadherin lacks a calcium-binding motif between EC2 and EC3 (9) and our crystal structure has confirmed that hypothesis experimentally. Crystal structures of cadherins which possess calcium-free linker have been reported previously and the biological significance of the calcium-free linker has been discussed (19, 20). 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. We performed SV-AUC using LI-cadherin EC1-5 and obtained a K_D value of 22.8 μM **(Fig. S2)**. 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.

Calcium-free linker

We first investigated the calcium-free linker between EC2 and EC3. Classical cadherins

generally adopt a crescent-like shape (17, 21). 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.

Generally, three calcium ions bound to the linker between each EC domain confer rigidity to the structure (11). In fact, previous study on calcium-free linker of cadherin has shown that the linker showed some flexibility (20). To compare the rigidity of the canonical linker with three calcium ions and the calcium-free linker in LI-cadherin, we performed MD simulations. In addition to the monomeric states, we also used the structure of the EC1-4 homodimer as the initial structure of the simulations. After confirming the convergence of the simulations by calculating RMSD values of Cα atoms **(Fig. S3, see Experimental Procedures for the details)**, 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. 4A, B)**. The EC3 domain in the monomer conformation exhibited the largest RMSD. The RMSD values of EC3 in the homodimer 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. S4A-C)**. These results showed that the calcium-free linker between EC2 and EC3 is more flexible than the canonical linker **(Movie 1, 2)**.

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. S5)**. The Asp and Glu residues of the DXE motif in LI-cadherin dimer EC1 and EC3 coordinate with calcium ions **(Fig. S6A, B)** and was maintained throughout the simulation **(Fig. S6C~J)**. The α-helix in EC2 might compensate for the absence of calcium by conferring some rigidity to the molecule.

Interaction analysis of EC1-4 homodimer

To validate if LI-cadherin-dependent cell adhesion is mediated by the formation of homodimer observed in the crystal structure, it was necessary to find a mutant which exhibits decreased dimerization tendency. First, we analyzed the interaction between two EC1-4 molecules in the crystal structure using the PISA server **(Table S3)** (22). 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. 5)**. The dimerization interface area was 1,253.3 $Å^2$ and the number of hydrogen bonds (distance between heavy atoms <

3.5 Å) was seven. Based on the analysis of these interactions, we conducted site-directed mutagenesis to assess the contribution of each residue 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. To quickly find the mutant with weaker homodimerization propensity, SEC-MALS was employed.

We injected EC1-4 WT or each mutant at 100 μM in the chromatographic column. Analysis of the molecular weight (MW) showed that the MW of F224A was the smallest among all the mutations evaluated, and also including WT **(Fig. 6A and Table 2)**. The same observation was made when the samples were injected at 50 μM **(Fig. S7A)**. It was also revealed that among the 12 samples analyzed, the elution volume of F224A was the largest **(Fig. 6B and Fig. S7B)**. In summary, the mutational study using SEC-MALS suggested that the mutation F224A was the most significant inhibiting homodimerization of EC1-4.

We must note that the samples eluted as a single peak, corresponding to a fast equilibrium between monomers and dimers as reported in a previous study employing other cadherins (23). Although the samples were injected at 100 μM, they eluted at ~ 4 μM since SEC will dilute the samples as they advance through the column. Considering that the K_D of dimerization of EC1-4 WT determined by AUC was 39.8 μM, at a protein concentration of 4 μM, the largest fraction of the eluted sample should be monomer. This explains why the MW of the WT sample was smaller than the MW of the homodimer (99.6 kDa), and why the differences in MW among the constructs were small. However, we assume that the decrease of MW and the increase of elution volume indicate the decrease of the proportion of homodimer in the eluted sample, indicating a smaller dimerization tendency caused by the mutations introduced in the protein.

Contribution of Phe224 to dimerization

Although F224 does not seem to form extensive specific interactions with the partner molecule of LI-cadherin in our crystal structure **(Fig. S8)**, its buried area upon dimerization was calculated to be 94% by the PISA server, engaging in Van der Waals interactions with other residues of EC1-4. To understand the role of Phe224 in dimerization of LI-cadherin, we conducted MD simulations of EC1-4 WT and EC1-4 F224A in the monomeric states, respectively. We first calculated the intramolecular 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. 7A, Fig. S9 and Movie 3, 4)**. This movement suggests that the side chain of Phe224 forms intramolecular interaction and is stabilized inside the pocket. Superposition of EC2 (chain A) in the crystal structure of EC1-4 and EC2 during the simulation of EC1-4 F224A monomer suggests that the large movement of the

loop including Ala224 would cause steric hindrance and would inhibit dimerization **(Fig. 7B)**.

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. 7C)**. These results suggested 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 restricting the movement of the residues around Phe224 and thus preventing the steric hindrance by the large movement as observed by MD simulations. DSC measurements showed that some of other mutants have lower thermal stability than wild type **(Table 2 and Fig. S10)**. However, because T_{M1} of F224A is the 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.

Functional analysis of LI-cadherin on cells

To investigate if the disruption of the formation of EC1-4 homodimer 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. S11)**. We conducted cell aggregation assays and compared the cell adhesion ability of cells expressing each construct and mock cells (non-transfected Flp-In CHO) in the presence of calcium or in the presence of EDTA. 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. 8A-C)**. From this result, it was revealed that F224 was crucial for LI-cadherin-dependent cell adhesion and the formation of EC1-4 homodimer in the cellular environment was indicated.

Difference of LI-cadherin and classical cadherin

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 and Fig. S11)**. Importantly, neither EC1-5 nor EC3-7 expressing cells showed cell aggregation ability in the presence of $CaCl_2$ **(Fig. 9)**.

EC1-5 expressing cells did not aggregate, suggesting that effective dimerization requires full-length protein. The EC1-4 homodimer observed by X-ray crystallography and detected by AUC cannot be replicated by the EC1-5 construct in a cellular

environment, suggesting that the overhang EC1 domain in the dimer belonging to one cell collides with the membrane of the opposing cell (steric hindrance) **(Fig. S12A)**. It is also possible that inappropriate orientation of the approaching LI-cadherin molecules would also contribute to the inability of EC1-5 to dimerize **(Fig. S12B)**. An alternative possibility is that the weaker dimerization of EC1-2 (detected by AUC) cannot maintain cell adhesion due to the mobility of the Ca^{2+}-free linker between EC2 and EC3. Contrary to the canonical Ca^{2+}-bound linker, such as the linker between EC1 and EC2, the linker between EC2 and EC3 in LI-cadherin does not possess a Ca^{2+}. The lack of Ca^{2+} resulted in greater mobility when EC1-4 homodimer observed by crystal structure **(Fig. 3)** was not formed. The combination of low dimerization affinity and high mobility likely explain the absence of EC1-2 driven cell adhesion **(Fig. S12C, 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 homodimer, which showed that Trp239 was buried in its own hydrophobic pocket and was not participating in homodimerization **(Fig. S13)**, 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. S14)**. This result excluded the possibility of nonsymmetrical interaction of the domains (e.g. EC1-2 and EC3-4, EC1-2 and EC6-7, etc.).

Discussion

Here, we show the homodimer architecture of LI-cadherin EC1-4 and the flexibility of Ca^{2+}-free linker in LI-cadherin monomer for the first time. The X-dimer or the strand-swap dimer formed by classical cadherins do not seem effective to drive LI-cadherin-dependent cell adhesion, as these dimers would lead to large movements at the Ca^{2+}-free linker even if the dimer was formed. We assume that the unique architecture of LI-cadherin EC1-4 homodimer was necessary to restrict the movement of Ca^{2+}-free linker to maintain LI-cadherin-dependent cell adhesion **(Fig. S15)**.

Several differences between LI-cadherin and E-cadherin might explain the reason for the existence of non-canonical Ca^{2+}-free linker. Both LI-cadherin and E-cadherin are expressed on normal intestine cells, however, their expression sites are different. LI-cadherin is expressed at intercellular cleft and is excluded from adherence junction (7), where E-cadherin is expressed (24). Even though LI-cadherin is excluded from adherence junction,

trans-interaction of LI-cadherin is necessary to maintain water transport through intercellular cleft of intestine cells (8). Clustering on cell membrane might also be different. Classical cadherins including E-cadherin are considered to form cluster on cell membrane to achieve cell adhesion (17). Lateral interaction interface of these cadherins was estimated from the crystal lattices. In contrast, we did not observe any crystal packing which suggests lateral interaction in our crystal structure. Indeed, our crystal structure shows that N-terminal sugar chains are extended toward the opposite side of the homodimer interface, and this suggests that each homodimer does not participate in cis-interaction. Considering that the interface area of the X-dimer and strand-swap dimer are much smaller than that of LI-cadherin EC1-4 dimer, we speculate that LI-cadherin form homodimers with a broader interface to be able to maintain trans-interaction without formation of clusters on the cell membrane.

Expression of LI-cadherin is also observed on various cancer cells such as gastric adenocarcinoma, colorectal cancer cells and pancreatic cancer cells (4, 25, 26). The roles of LI-cadherin on cancer cells have been discussed previously. For example, it was shown that inoculation of LI-cadherin gene (CDH17)-silenced cells in nude mice inhibited the progression of colorectal cancer (27). In case of gastric cancer, the size of LI-cadherin-positive tumor was significantly larger than that of LI-cadherin-negative tumor (28). Considering that loss of cell adhesion ability by the downregulation of E-cadherin by epithelial mesenchymal transition (EMT) is often observed in cancer cells (29, 30), the fact that LI-cadherin is upregulated in various types of cancer cells suggest that LI-cadherin acts differently with E-cadherin on cancer cells. The unique architecture of the LI-cadherin homodimer and the absence of interactions with intracellular (cytoplasmic) proteins (31) suggest a distinctive role of LI-cadherin in cancer cells with respect to that of classical cadherins.

In summary, our study shows the novel characteristics of LI-cadherin at the molecular level. Our results suggest that molecules targeting interface of LI-cadherin homodimer abrogate the LI-cadherin-dependent cell adhesion. On the other hand, we estimate that molecules which restrict the movement of Ca^{2+}-free linker might strengthen LI-cadherin-dependent cell adhesion by stabilizing LI-cadherin homodimer.

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 (Cytiva) 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.

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) (32) 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 (33) 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) (34). 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) (35).

Solution structure analysis using SAXS

All measurements were performed at beamline BL-10C (36) of the Photon Factory (Tsukuba, Japan). The experimental procedure is described previously (18). Concentrations of EC3-4 was 157 μM. Data were collected using a PILATUS3 2M (Dectris). A wavelength was 1.488 Å with a camera distance 101 cm. 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 four measurements 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. The calculation of the theoretical curves of SAXS and χ^2 values were performed using FoXS server (37, 38).

MD simulation

Molecular dynamics simulations of LI-cadherin were performed using GROMACS 2016.3 (39) with the CHARMM36m force field (40). A whole crystal structure of EC1-4 homodimer, EC1-4 monomer form, EC1-4 F224A monomer form and EC3-4 monomer form was used as the initial structure of the simulations, respectively. EC1-4 and EC3-4 of chain A was extracted from EC1-4 homodimer crystal structure to generate EC1-4 monomer form and EC3-4 monomer form, respectively. Sugar chains were removed from the original crystal structure. Missing residues were modelled by MODELLER 9.18 (41). Solvation of the structures were performed with TIP3P water (42) 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 (43). 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 (43, 44). 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 (45). Covalent bonds involving hydrogen atoms were constrained by the LINCS algorithm (46). A snapshot was saved every 10 ps. All trajectories were analyzed using GROMACS tools. RMSD, dihedral angles, distances between two atoms and clustering were computed by rms, 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 were excluded from the calculation of RMSD of EC1 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.

Generation of EC3-4_plus

MD simulation of the EC3-4 monomer was performed for 220 ns. The trajectories from 20 ns to 220 ns were clustered using the 'cluster' tool of GROMACS. The structure which exhibited the smallest average RMSD from all other structures of the largest cluster was termed EC3-4_plus and used for the purpose of comparison with the data in solution (SAXS).

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 (47) of the CCP4 suite (48). The structure of the WT protein was determined by the molecular replacement method using the coordinates of P-cadherin (PDB entry code 4ZMY) (49) with the program PHASER (50). The models were refined with the programs REFMAC5 (51) and built manually with COOT (52). Validation was carried out with PROCHECK (53). Data collection and structure refinement statistics are given in Table 1. UCSF Chimera was used to render all of the molecular graphics (54).

Site-directed mutagenesis

Introduction of mutation to plasmid was performed as described previously (55).

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 (Cytiva) 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). Concentration at the end of the chromatographic column was measured based on the UV absorbance. The protein conjugate method was employed for the analysis as sugar chains were bound to LI-cadherin. All detectors were calibrated using bovine serum albumin (BSA) (Sigma-Aldrich).

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 (Cytiva). 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 (Cytiva) 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 (56, 57). 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.

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. The plate was incubated on ice for more than 20 minutes. Images of the cells were taken using EVOS® XL Core Imaging System (ThermoFisher Scientific) if necessary. After that, cells were injected to MFI.

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.) and a Grant-in-Aid for Scientific Research (B) 20H02531 (K.T.) from Japan Society for the Promotion of Science, a Grant-in-Aid for Scientific Research on Innovative Areas 19H05760 and 19H05766 (K.T.) from Ministry of Education, Culture, Sports, Science and Technology, and a Grant-in-Aid for JSPS fellows 18J22330 (A.Y.) from Japan Society for the Promotion of Science. 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. 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

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 Ca2+-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. 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
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 (80-.).* **296**, 1308–1313
18. 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
19. 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
20. 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
21. 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
22. Krissinel, E., and Henrick, K. (2007) Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **372**, 774–797
23. 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
24. Boller, K., Vestweber, D., and Kemler, R. (1985) Cell-adhesion molecule uvomorulin is localized in the intermediate junctions of adult intestinal epithelial cells. *J. Cell Biol.* **100**, 327–332
25. Grötzinger, C., Kneifel, J., Patschan, D., Schnoy, N., Anagnostopoulos, I., Faiss, S., Tauber, R., Wiedenmann, B., and Geßner, R. (2001) LI-cadherin: A marker of gastric metaplasia and

neoplasia. *Gut*. **49**, 73–81

26. Liu, X., Huang, Y., Yuan, H., Qi, X., Manjunath, Y., Avella, D., Kaifi, J. T., Miao, Y., Li, M., Jiang, K., and Li, G. (2019) Disruption of oncogenic liver-intestine cadherin (CDH17) drives apoptotic pancreatic cancer death. *Cancer Lett.* **454**, 204–214
27. Bartolomé, R. A., Barderas, R., Torres, S., Fernandez-Aceñero, M. J., Mendes, M., García-Foncillas, J., Lopez-Lucendo, M., and Casal, J. I. (2014) Cadherin-17 interacts with α2β1 integrin to regulate cell proliferation and adhesion in colorectal cancer cells causing liver metastasis. *Oncogene*. **33**, 1658–1669
28. Wang, J., Yu, J. C., Kang, W. M., Wang, W. Z., Liu, Y. Q., and Gu, P. (2012) The predictive effect of cadherin-17 on lymph node micrometastasis in pN0 gastric cancer. *Ann. Surg. Oncol.* **19**, 1529–1534
29. Huang, R. Y. J., Guilford, P., and Thiery, J. P. (2012) Early events in cell adhesion and polarity during epithelialmesenchymal transition. *J. Cell Sci.* **125**, 4417–4422
30. Lamouille, S., Xu, J., and Derynck, R. (2014) Molecular mechanisms of epithelial–mesenchymal transition. *Nat. Rev. Mol. Cell Biol.* **15**, 178–196
31. 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
32. Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78**, 1606–1619
33. 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.*
34. Brautigam, C. A. (2015) Chapter Five - Calculations and Publication-Quality Illustrations for Analytical Ultracentrifugation Data. in *Methods in Enzymology* (Cole, J. L. B. T.-M. in E. ed), pp. 109–133, Academic Press, **562**, 109–133
35. Schuck, P. (2003) On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal. Biochem.* **320**, 104–124
36. 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
37. 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
38. 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

39. 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
40. 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
41. 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
42. 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
43. 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
44. 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
45. 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
46. 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
47. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 72–82
48. 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
49. 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
50. 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

51. Murshudov, G. N. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.* **53**, 240–255
52. 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
53. Laskowski, R. A. (1993) PROCHECK-a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291
54. 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
55. 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
56. 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
57. 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

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 Ala scanning

Sample	SEC-MALS		DSC		
	Concentration (μM)	MW[1] (kDa)	Concentration (μM)	T_{M1} (°C)	T_{M2} (°C)
WT	100	53.7	7.9	60.1 ± 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	50.7			
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	51.6			
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	52.4	7.1	59.5 ± 0.0	N.D.[3]
	50	N.D.[3]			
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	52.7	7.8	59.6 ± 0.0	N.D.[3]
	50	N.D.[3]			

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.

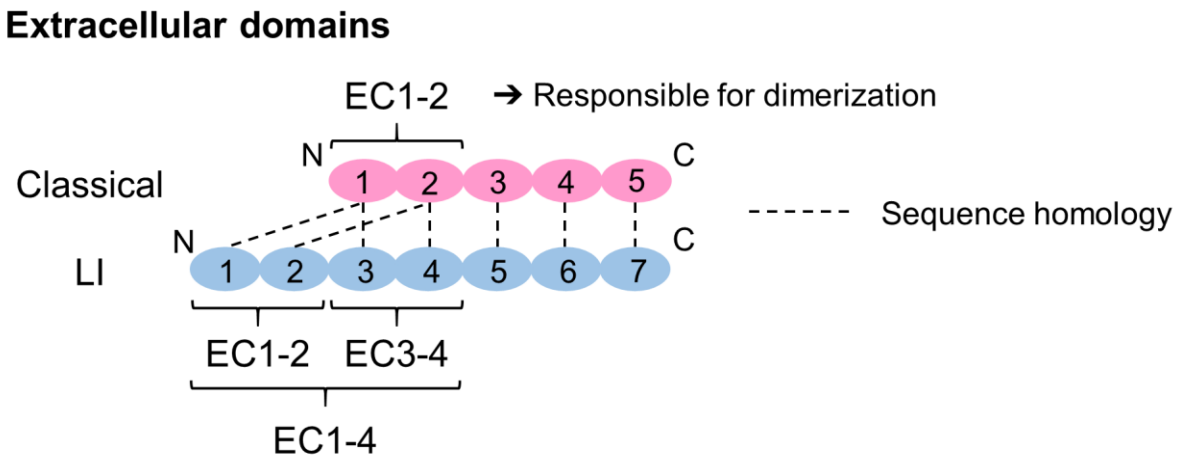

Figure 1. Schematic view of extracellular cadherin (EC) domains of classical cadherin and LI-cadherin. Domains connected by dotted lines have sequence homology. EC1-4, EC1-2 and EC3-4, which were used for the experiments are indicated by parentheses.

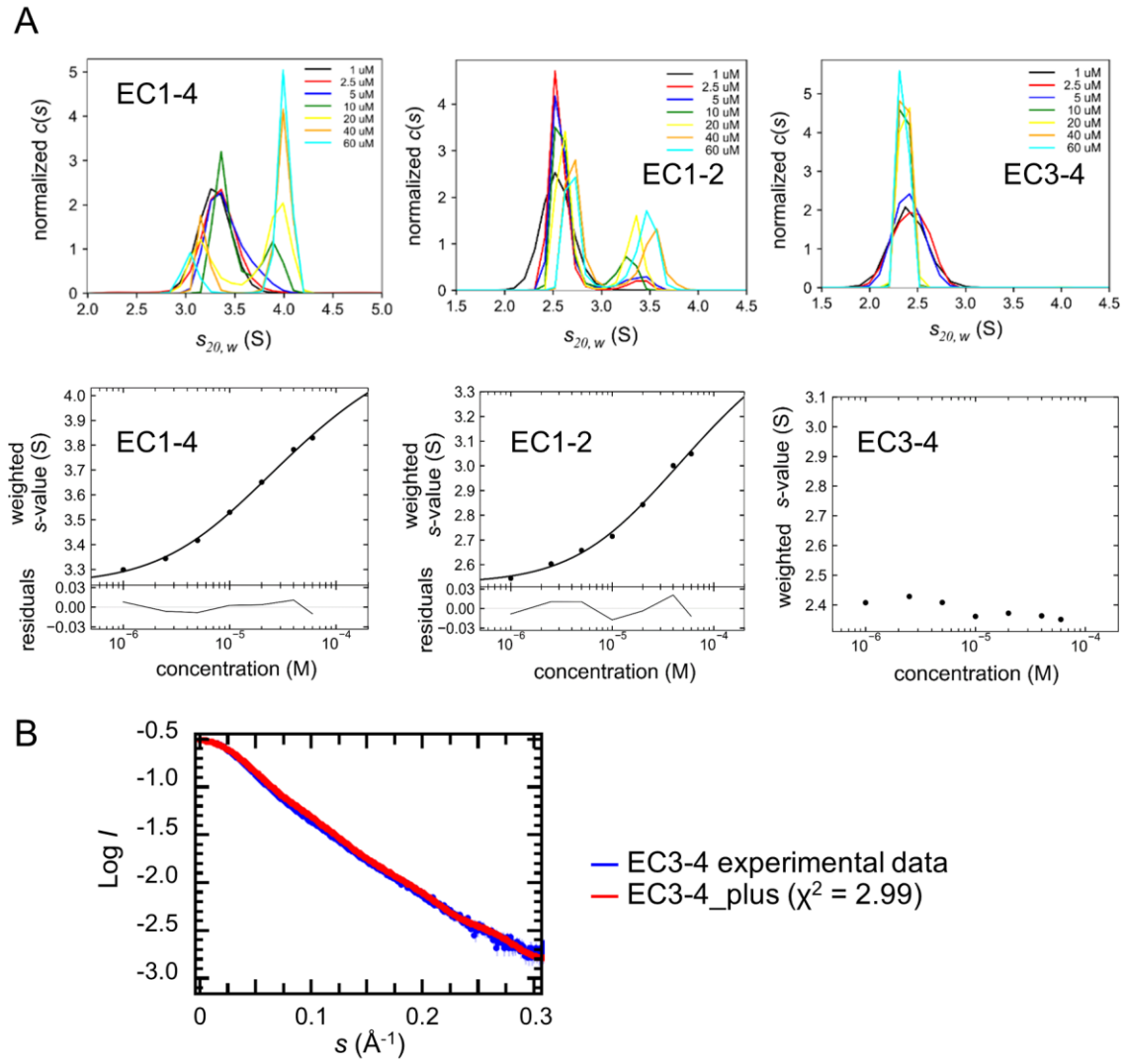

Figure 2. Analysis on dimerization state of LI-cadherin EC1-4, EC1-2 and EC3-4. **A.** Sedimentation plot of SV-AUC. Dimerization of EC1-4 and EC1-2 was confirmed. K_D of EC1-4 and EC1-2 homodimer was 39.8 μM and 75.0 μM, respectively. **B.** Scattering curve of SAXS and theoretical curve of EC3-4 calculated from modified crystal structure. Method to produce modified crystal structure is explained in experimental procedures and supplementary information.

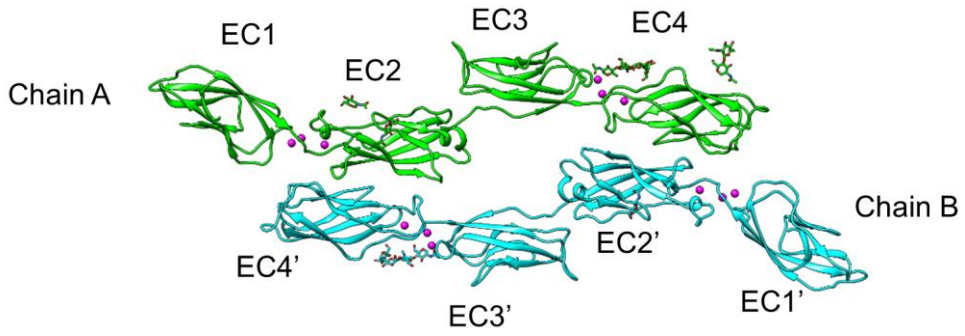

Figure 3. Crystal structure of EC1-4 homodimer. Calcium ions are depicted in magenta. No calcium ions were observed between domains EC2 and EC3 in either molecule. Four partial N-glycans were modeled in chain A (light green) and two in chain B (cyan) (the amino acid sequence of EC1-4 is given in Table S1).

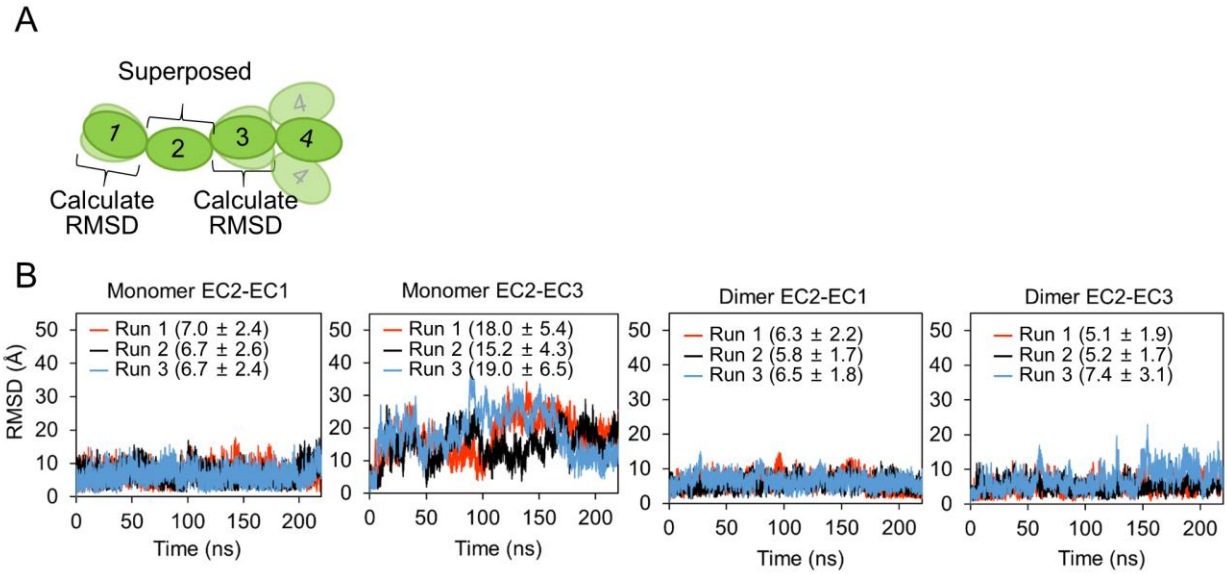

Figure 4. Computational analysis of the flexibility of calcium-free linker. **A.** Schematic view of how RMSD values were calculated. **B.** RMSD values of EC1 Cα or EC3 Cα against EC2 Cα. Chain A of EC1-4 dimer structure was employed as the initial structure. **C.** RMSD values of EC1 Cα or EC3 Cα in chain A of the dimer structure against EC2 Cα in the chain A. RMSD values and standard deviations are shown in parenthesis in angstrom unit.

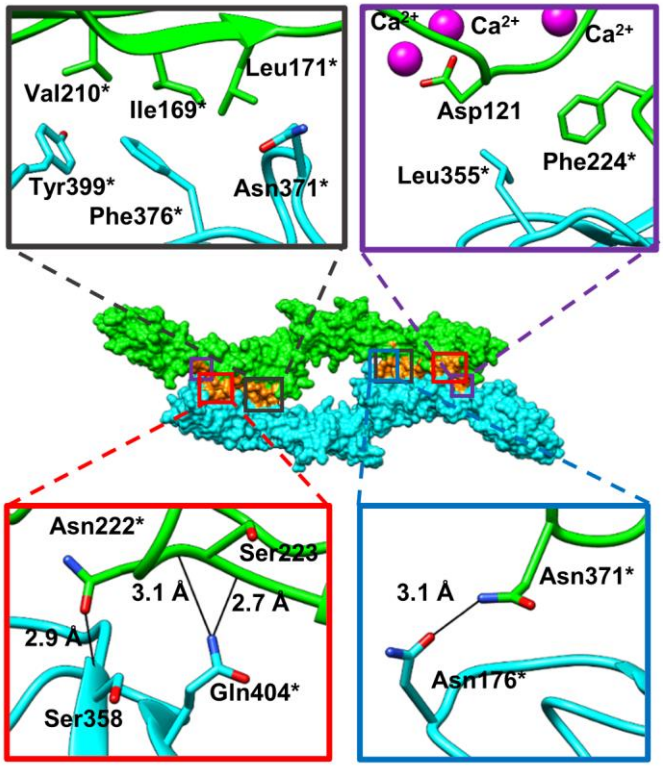

Figure 5. Residues involved in the intermolecular interaction in EC1-4 homodimer crystal structure. The non-polar interaction residues are shown in black and purple rectangles (top panels). Residues involved in hydrogen bonds (black solid lines) are shown within the red and blue rectangles (bottom panel). Residues indicated with an asterisk were individually mutated to Ala to evaluate their contribution to dimerization.

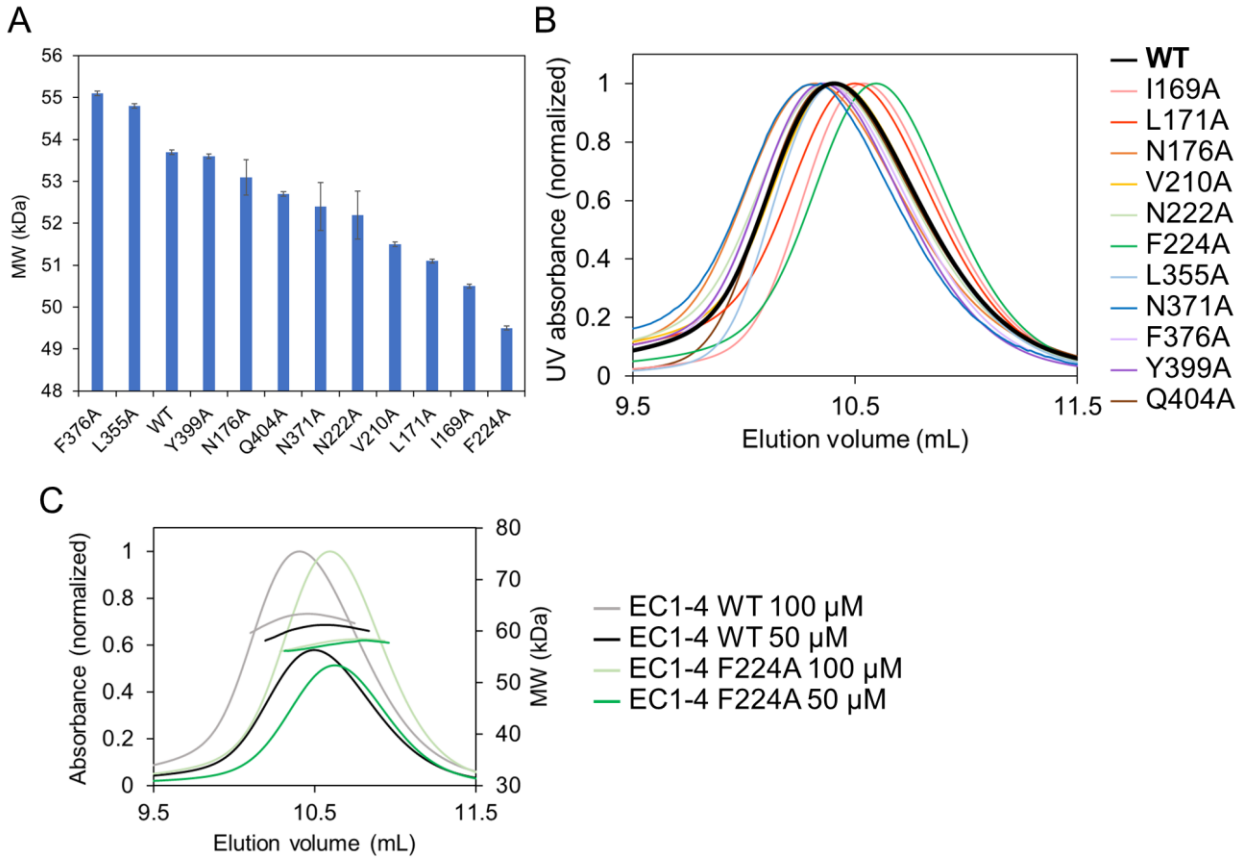

Figure 6. Mutagenesis analysis by SEC-MALS. **A.** Molecular weight measured by MALS. F224A exhibited the smallest molecular weight among all constructs tested. The samples were injected at 100 μM. Error bars indicate experimental uncertainties. **B.** SEC chromatograms obtained using SEC-MALS. Protein was injected at 100 μM. Chromatogram of WT and F224A are indicated in black (bold line) and green, respectively. Elution volume of the peak top of F224A was the largest among all constructs. **C.** SEC chromatogram and MW plots of EC1-4 WT and F224A. Graphs of other mutants are shown in Fig. S7C.

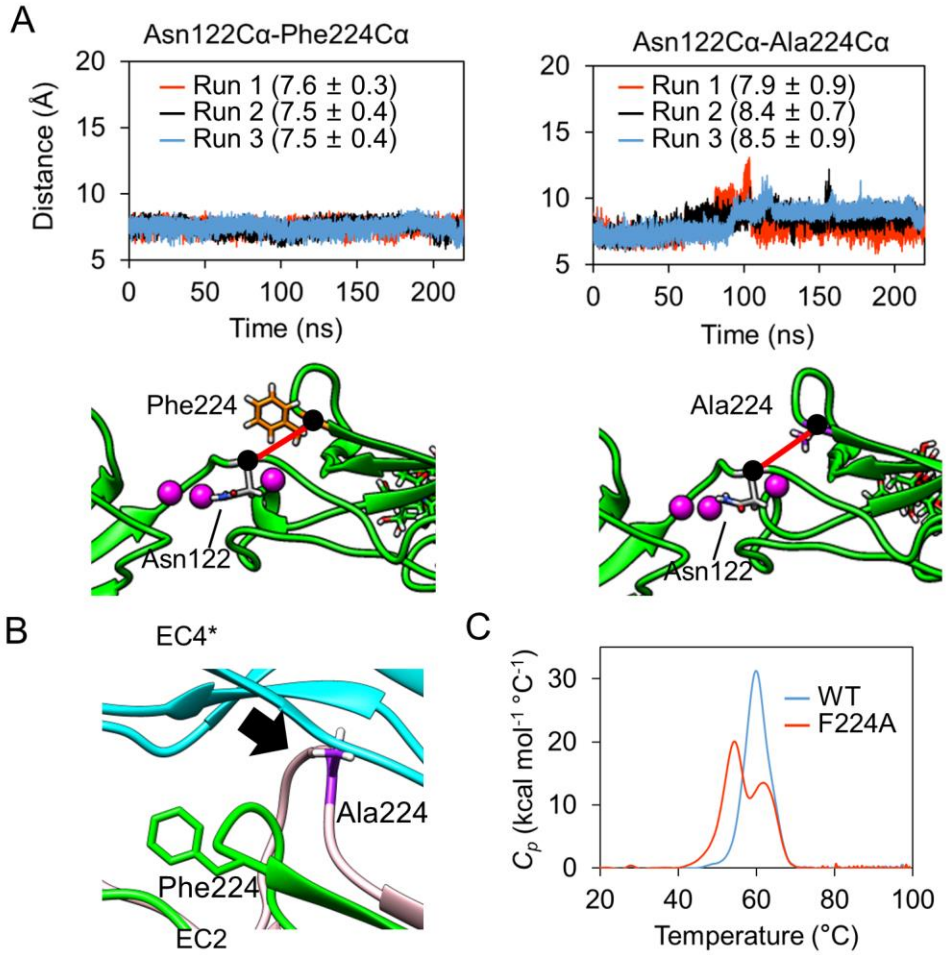

Figure 7. Mechanism of dimerization facilitated by Phe224. **A.** The distance between Phe224 (orange) Cα or Ala224 (purple) Cα and Asn122 (grey) Cα was evaluated by MD simulations. The Cα atoms are indicated by black circles. The distance calculated by the simulations is indicated with red line. Time course is shown on the portion of the panel at the upper part of each structure. Each MD simulation 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.** Structure alignment of EC2 (chain A) in EC1-4 homodimer crystal structure and EC2 during the MD simulation of EC1-4F224A monomer. A snapshot of 103.61 ns in Run 1 was chosen as it showed the largest distance between Asn122 and Ala224. Ala224 is indicated in purple. The loop indicated with the black arrow would cause steric hindrance towards the formation of the homodimer. **C.** 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.

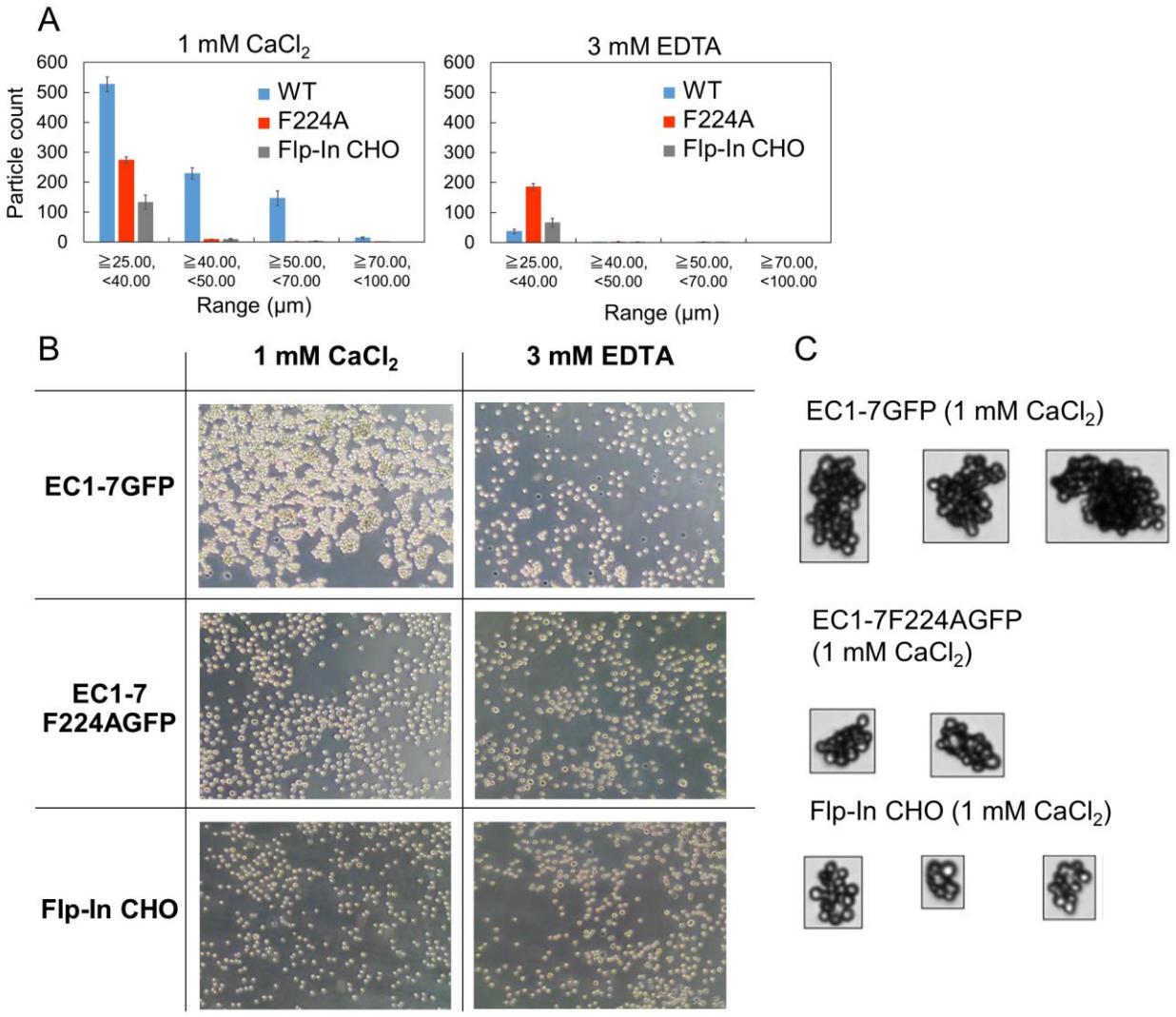

Figure 8. Cell aggregation assay. **A.** Size distribution of cell aggregates determined by MFI. Particles that were 25 μm or larger were regarded as cell aggregates. Only EC1-7 WT expressing cells in the existence of 1 mM $CaCl_2$ showed significant number of cell aggregates that were 40 μm or larger. Data show the mean ± SEM of four measurements. **B.** Microscopic images of cell aggregates taken after adding 4% PFA and incubating the plate on ice for 20 minutes. **C.** Images of cell aggregates taken by MFI. Cell aggregates belonging to the largest size population of each construct obtained in the presence of 1 mM $CaCl_2$ (70~100 μm for EC1-7GFP, 50~70 μm for EC1-7F224AGFP and 40~70 μm for Flp-In CHO) are shown.

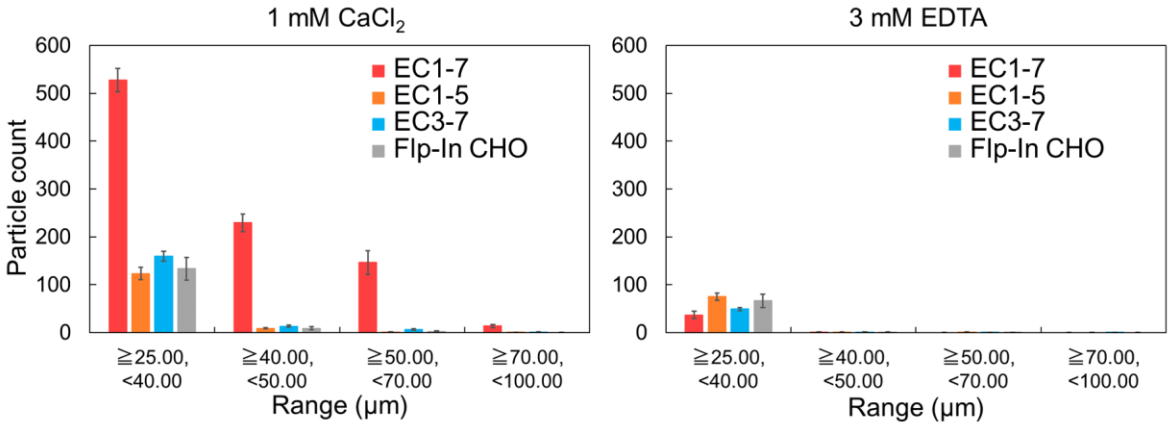

Figure 9. Cell adhesion mediated by short constructs. Cell aggregation assay using EC1-5GFP and EC3-7GFP. EC1-7GFP and Flp-In CHO (mock cells) were used as positive and negative control, respectively. Particles that were 25 μm or larger were considered as cell aggregates. The number of cell aggregates of both EC1-5GFP and EC3-7GFP in the presence or absence of Ca^{2+} were determined. Data show mean ± SEM of four measurements.