Supporting Information

Tariq et al. 10.1073/pnas.1420508112

SI Materials and Methods

Cloning of mDsg2 Ectodomain. A construct containing the signal peptide, propeptide, and full ectodomain (EC1–EC5) of mDsg2 (residues 1–614 from the precursor sequence, SwissProt entry O55111), followed in-frame by a C-terminal hexahistidine tag, was produced for expression in mammalian cells. The construct named pCAGG-mDsg2WED was generated by replacing human E-cadherin from the pC-E-cadHA construct (kind gift from Masayuki Ozawa, Kagoshima, Japan) with a PCR-generated mouse Dsg2 full ectodomain (GenBank accession no. AB072269). The sense primer was 5′-gacaaGCGGCCGCcggcggatcgaggcg***ATG***gcgcg-gagc, and the antisense primer was 5′-gacaaGATATC***TCAgtgatggt-ggtgatggtg***cccgacgtagttgtcatactgtgc. Restriction sites, NotI in sense primer and EcoRV in antisense primer, are shown in uppercase, and start codon ATG and stop codon TCA are in bold italic.

Protein Expression and Purification. The pCAGG-mDsg2WED construct was transfected into CHO cells for expression in serum-free CHO CD3 medium (Sigma) in suspension culture at 37 °C at 120 rpm. The expressed protein was confirmed by peptide mass fingerprinting and N-terminal sequencing that showed the first five amino acids were AWITAP expected N-terminal sequence of correctly processed mDsg2. Supernatant containing the secreted protein was harvested by centrifugation at 16,260 × *g*, 10 min, 4 °C, concentrated and buffer-exchanged into 20 mM Tris·Base at pH 8.0, 0.5 M NaCl, by tangential flow filtration (Pall). Purification was done using Ni-NTA agarose beads (Qiagen) with 20 mM Tris·Base at pH 8.0, 0.5 M NaCl, and eluted with imidazole followed by SEC, using a Superdex-200 (GE Healthcare) in 20 mM Tris·Base at pH 8.0, 0.5 M NaCl, and 1 mM EDTA with or without 5 mM $CaCl_2$ (SEC buffer).

Multiangle Laser Light Scattering. Samples of mDsg2 were injected onto a Superdex-200 equilibrated with SEC buffer and elution samples passed through a Wyatt Helios 18-angle laser photometer, coupled to a Wyatt Optilab rEX refractive index detector. Data were analyzed using the Astra 6.1 software. The addition of calcium results in an earlier elution volume in the SEC column (0.7 mL shift) compared with the EDTA sample (Fig. S1), indicating that the molecule becomes more elongated and elutes faster. Higher concentrations of $CaCl_2$, up to 10 mM, produced no further effect on the elution profile, suggesting the system was already saturated for calcium binding at 2mM Ca^{2+}.

Analytical Ultracentrifugation. mDsg2 (0.1–0.5 mg/mL) in the SEC buffer was used for the experiments. Sedimentation coefficients were determined by sedimentation velocity experiments using the Optima XL-A ultracentrifuge (Beckman Instruments). The experiments were performed using double-sector cells and a rotor speed of 147,000 × *g*, taking 200 scans at 11/2-min intervals at a wavelength of 230 or 280 nm at 20 °C. Data were analyzed with Sedfit v8.7 (1), and hydrodynamic radius (R_h) and frictional ratio (f/f_0) were calculated with Sednterp (2).

Small-Angle X-ray Scattering Analysis. SAXS measurements of mDsg2 (5 mg/mL) in the SEC buffer were collected on the I22 beam line at Diamond Light Source. The beam was focused onto the photon-counting detector, placed at a distance of 2.5 m from the sample cell, using X-ray wavelength of 0.1 nm. The covered range of momentum transfer was $0.014 < q < 0.48$ Å^{-1} ($q = 4\pi\sin\theta/\lambda$). The q range was calibrated using silver behenate powder based on diffraction spacings of 58.38 Å. The data were normalized to the intensity of the incident beam and spherically averaged using an in-house program. Data were collected in 120 successive 1-s frames. Data processing, Guinier analysis, and estimation of the radius or gyration (R_g) were done in PRIMUS (3). GNOM (4) was used to compute the pairwise intraparticle distance distribution function $p(r)$ and D_{max}. Particle shapes were restored ab initio in DAMMIN (5). Twenty simulations were performed, and the outputs were averaged and filtered using DAMAVER (6) to produce a final envelope with a normal spatial discrepancy value of 0.69.

Generation of Molecular Models. The MALLS analysis and the mobility shift observed after PNGase treatment suggested that mDsg2 contains at least two *N*-glycosylated sites. Four *N*-glycosylation sites have been predicted for mDsg2 (Fig. S1*A*), and one site (N413) has been identified by mass spectrometry (7), but the remaining sites (N63, N260, and N465) have not yet been confirmed. Only two of them are fully accessible (N260 in EC3 and N413 in EC4), and therefore these *N*-glycans were built into the initial EC molecular models. The conformational space of the sugars was explored with restrained TAMD, as implemented in CNS (8), followed by comparison with the experimental SAXS profile with FoXS to choose the best conformations. The program Modeler, version 9.8 (9), was used to build a homology model of the mDsg2 EC1–EC5 ectodomain, using the hDsg2 model as a template (10). Then the five individual EC1–EC5 domains were used in EOM to generate the pool of initial models.

Alternative approaches were used to generate a subset of conformers to model the EC of mDsg2 that will match the SAXS and biophysical data. The initial pool of 10,000 random models from EOM was used to calculate hydrodynamic parameters with SoMo (11) for each model, including gyration radius (R_g), hydrodynamic radius (R_h), sedimentation coefficient (s), maximal linear distance (D_{max}), and corrected intrinsic viscosities and frictional ratios (f/f_o). Experimental hydrodynamic data from MALLS and AUC were used to select a subset of models that fit the range of s values. Theoretical scattering curves were calculated for each selected model and compared with the experimental SAXS profiles with GAJOE (12) and FoXS (13), and χ^2_{free} were calculated with Scatter (14). In parallel, the full ectodomain model was subject to TAMD to generate 500 different models and subsequently filtered against the experimentally hydrodynamic parameters.

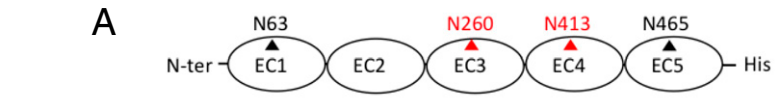
Modeling confirmed that the final selected models for all approaches were compatible with calcium coordination of at least the six calcium-binding sites between EC1–EC2 and EC2–EC3 and partially elsewhere, with weaker coordination (longer distances) at one of the EC3–EC4 Ca^{2+} and one of the EC4–EC5 Ca^{2+} ions, as described in Fig. 4. This means that although highly flexible, Dsg2 can still fulfill most of the Ca-binding.

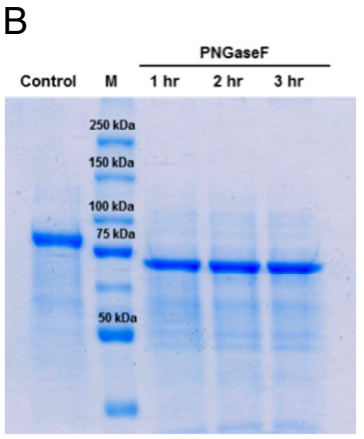
Electron Microscopy. Samples of mDsg2 were diluted to 20 μg/mL in Tris buffer containing 1 mM EDTA and $CaCl_2$ concentrations of 0, 2, and 5 mM. The samples were adsorbed to carbon-filmed 400-mesh copper grids, which were made hydrophilic immediately before use by treatment in a glow discharge vacuum unit with a current setting of 20 mA and exposure time of 20 s with a negative polarity. The grids were briefly stained with 1.25% uranyl acetate (pH 4.2), excess stain was wicked off, and they were finally air-dried. Low-dose images (<10 e/Å^2) were obtained on an FEI Tecnai 12 twin-electron microscope (FEI) operating at 120 keV. Images were mainly recorded at an instrumental

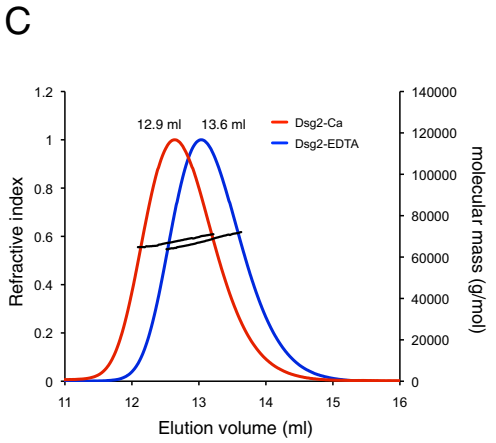
magnification of ×42,000 on a 2,000 × 2,000-pixel cooled CCD camera (TVIPS). The pixel size of the images was measured as 4.10 Å, using a negatively stained catalase grid as a magnification standard (Agar Scientific). Length measurements were made on the molecular images using ImageJ (15). Molecules were selected for measurement on the basis of having a clear separation from neighboring molecules and a distinct stain outline.

Molecular Fitting into the Desmosome Electron Tomography Maps and Generation of 3D Arrays. The DAMMIN model was used for docking in the ET map of native epidermal desmosomes (16) (EM DataBank entry 1374). The raw, unfiltered map shows three regions of contiguous density compatible with the dimensions of an entire mDsg2 ectodomain, plus three other regions with fragmented density. The DAMMIN model was docked into the six positions of the map compatible with entire mDsg2 ectodomains. Masks for the six docked models were generated, and sixfold averaging was applied, improving the quality of the original ET map, thus confirming the correct fitting. The six fitted models gave an approximate indication of a regular lattice. An idealized lattice model was then derived by fixing the central transdimer of mDsg2 models (as seen in Fig. 5) and applying the average translations along rows and columns. Images were produced with PyMol (17) and Chimera (18). See Movie S1 prepared with PyMol.

1. Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* 78(3):1606–1619.
2. Hayes D, Laue T, Philo J (1995) *Program Sednterp: Sedimentation Interpretation Program* (Alliance Protein Laboratories, Thousand Oaks, CA).
3. Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI (2003) PRIMUS: A Windows-PC based system for small-angle scattering data analysis. *J Appl Cryst* 36(5): 1277–1282.
4. Svergun DI (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Cryst* 25(4):495–503.
5. Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* 76(6):2879–2886.
6. Volkov VV, Svergun DI (2003) Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Cryst* 36(3):860–864.
7. Wollscheid B, et al. (2009) Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nat Biotechnol* 27(4):378–386.
8. Brunger AT (2007) Version 1.2 of the Crystallography and NMR system. *Nat Protoc* 2(11):2728–2733.
9. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234(3):779–815.
10. Garrod DR, Berika MY, Bardsley WF, Holmes D, Tabernero L (2005) Hyper-adhesion in desmosomes: Its regulation in wound healing and possible relationship to cadherin crystal structure. *J Cell Sci* 118(Pt 24):5743–5754.
11. Rai N, et al. (2005) SOMO (SOlution MOdeler) differences between X-Ray- and NMR-derived bead models suggest a role for side chain flexibility in protein hydrodynamics. *Structure* 13(5):723–734.
12. Bernadó P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc* 129(17):5656–5664.
13. Schneidman-Duhovny D, Hammel M, Sali A (2010) FoXS: A web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res* 38:W540–W544.
14. Rambo RP, Tainer JA (2013) Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* 496(7446):477–481.
15. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9(7):671–675.
16. Al-Amoudi A, Díez DC, Betts MJ, Frangakis AS (2007) The molecular architecture of cadherins in native epidermal desmosomes. *Nature* 450(7171):832–837.
17. The PyMOL Molecular Graphics System. Version 1.5.0.4. New York: Schrodinger, LLC.
18. Pettersen EF, et al. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605–1612.
19. Bohne-Lang A, von der Lieth CW. GlyProt: in silico glycosylation of proteins. *Nucleic Acids Res* 2005;33:W214–W219.
20. Vester-Christensen MB, et al. (2013) Mining the O-mannose glycoproteome reveals cadherins as major O-mannosylated glycoproteins. *Proc Natl Acad Sci USA* 110(52): 21018–21023.
21. Boggon TJ, et al. (2002) C-cadherin ectodomain structure and implications for cell adhesion mechanisms. *Science* 296(5571):1308–1313.
22. Harrison OJ, et al. (2011) The extracellular architecture of adherens junctions revealed by crystal structures of type I cadherins. *Structure* 19(2):244–256.
23. Larkin MA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21): 2947–2948.







D

	R_h (nm)	$S_{20,w}$	f/f_o	R_g (nm)	D_{max} (nm)
Dsg2-Ca experimental	4.6	3.49 ± 0.06	1.72	5.1 ± 0.30	17.5
Dsg2-EDTA experimental	4.1	3.91 ± 0.09	1.54	4.0 ± 0.15	14.0
Dsg2-Ca DAMMIN	4.7	3.4	1.8	4.9	17.2
Dsg2-EDTA DAMMIN	4.2	3.6	1.6	4.4	13.4
C-cadherin (1L3W)	4.6	3.2	1.8	6.5	22.0
E-cadherin (3Q2V)	4.5	3.4	1.7	6.3	21.4
N-cadherin (3Q2W)	4.5	3.4	1.7	6.4	22.2

Fig. S1. (*A*) Domain architecture of the recombinant mDsg2 ectodomain (residues 1–560 from the mature chain) containing five cadherin domains, EC1–EC5, followed by a C-terminal His_6-tag. Potential *N*-linked glycosylation sites (NetNGlyc 1.0 Server) are indicated as triangles. Sites predicted by GlyProt (19) to be exposed are shown in red. (*B*) SDS/PAGE with Coomassie staining of recombinant mDsg2 EC, before and after digestion with PNGase F. Lane M: standard molecular weight markers. Undigested mDsg2 (Control) migrates above 75 kDa. After treatment with PNGase F for 1–3 h, a shift is observed showing a band with lower molecular weight that corresponds to deglycosylated mDsg2. Incubations longer than 1 h have no further effect on migration. (*C*) SEC-MALLS profile of mDsg2-Ca and mDsg2-EDTA. The chromatogram shows refractive index (red and blue) and molecular mass (black) versus elution volume. The molecular mass measured in both cases is 68 kDa, but the elution volumes indicate a more compact structure in the absence of Ca^{2+} (blue). (*D*) Table showing the hydrodynamic and dimensional data for mDsg2 determined from AUC, MALLS, and SAXS, and parameters calculated with SoMo for the SAXS ab initio model and PDB structures of type 1 cadherins. R_h, hydrodynamic radius; $s_{20,w}$, sedimentation coefficient in water at 20 °C; f/f_0, frictional coefficient; R_g, radius of gyration; D_{max}, maximal linear dimension of the particles.

A

Sample	Measurement	Mean	SD	N
Dsg2-EDTA	Diameter (nm)	6.91	1.13	70
Dsg2-Ca	Contour length (nm)	19.2	2.62	55
	Width (nm)	3.22	0.53	55
	End-to-end length / contour length	0.84	0.13	55

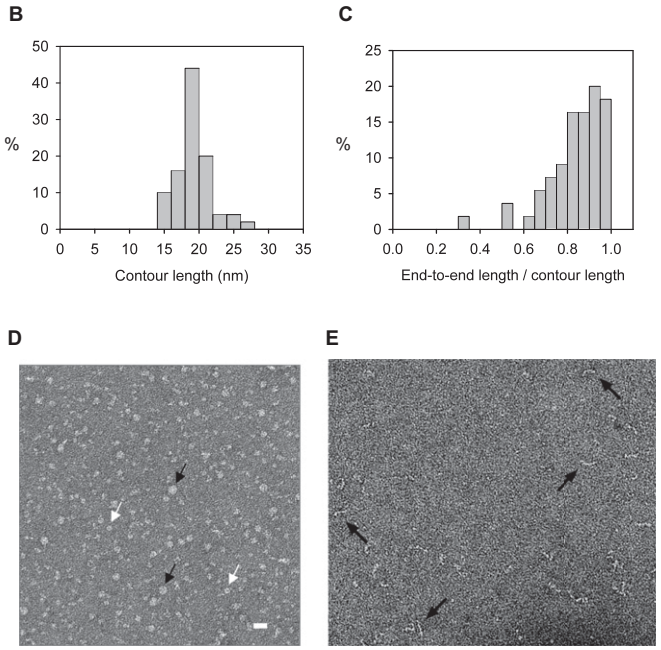


Fig. S2. (*A*) Table with molecular measurements on EM images of Dsg2-Ca and Dsg2-EDTA after negative staining. Contour length was measured as the total midline distance along the elongated molecule, and end-to-end length was measured as the linear distance between the two extreme ends of the molecule. (*B*) Histogram showing the frequency distribution of contour length measures for Dsg2-Ca. (*C*) The ratio between end-to-end and contour lengths gives an indication of the degree of elongation of a polymeric chain (with a value of 1 indicating a rod particle). The distribution of ratios illustrates the conformational variability of the mDsg2 ectodomain in calcium. A typical field of view is shown in *D* for DSG2-EDTA, where 97% of the molecules show a compact globular structure. Molecular aggregates also occur on the grid surface (examples indicated by black arrows), as well as the monomer species (examples indicated by white arrows). In contrast, flexible rod-like particles are observed in the presence of calcium (*E*). (Scale bar, 20 nm.)

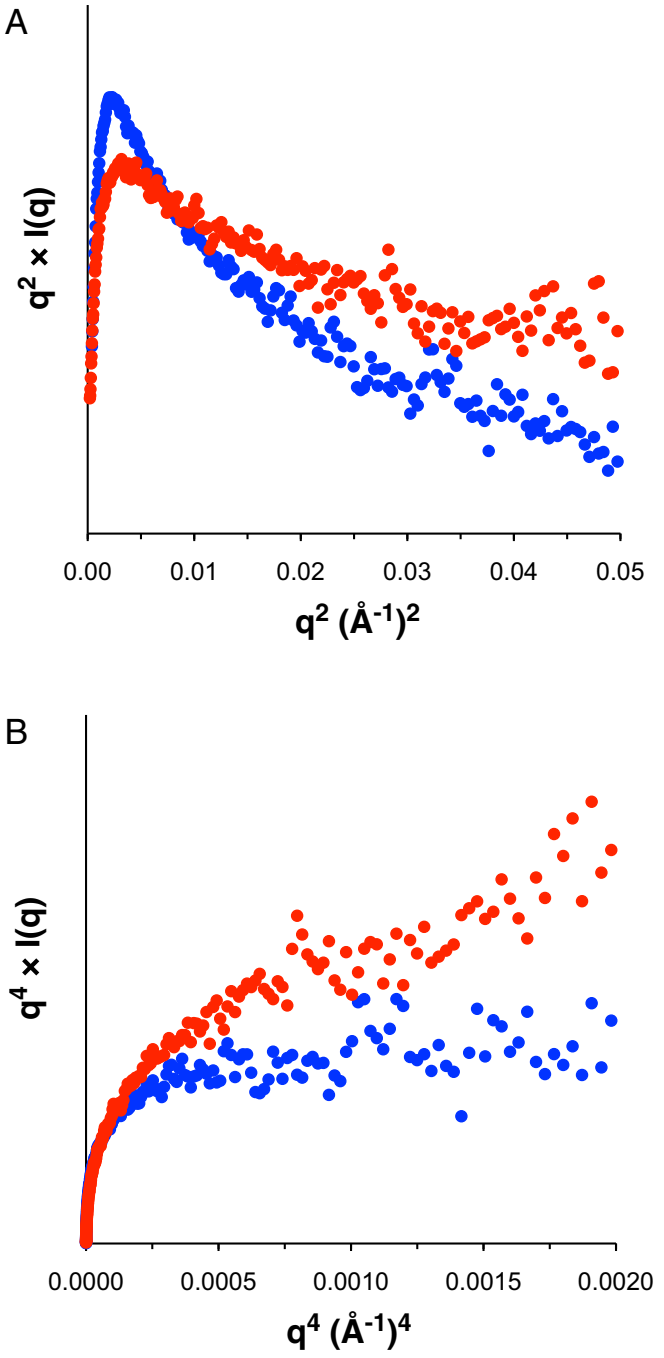


Fig. S3. Analysis of Dsg2 SAXS data. (*A*) Kratky-Debye plot in the absence (blue) and presence of Ca^{2+} (red). (*B*) Porod-Debye plot in the absence (blue) and presence of Ca^{2+} (red).

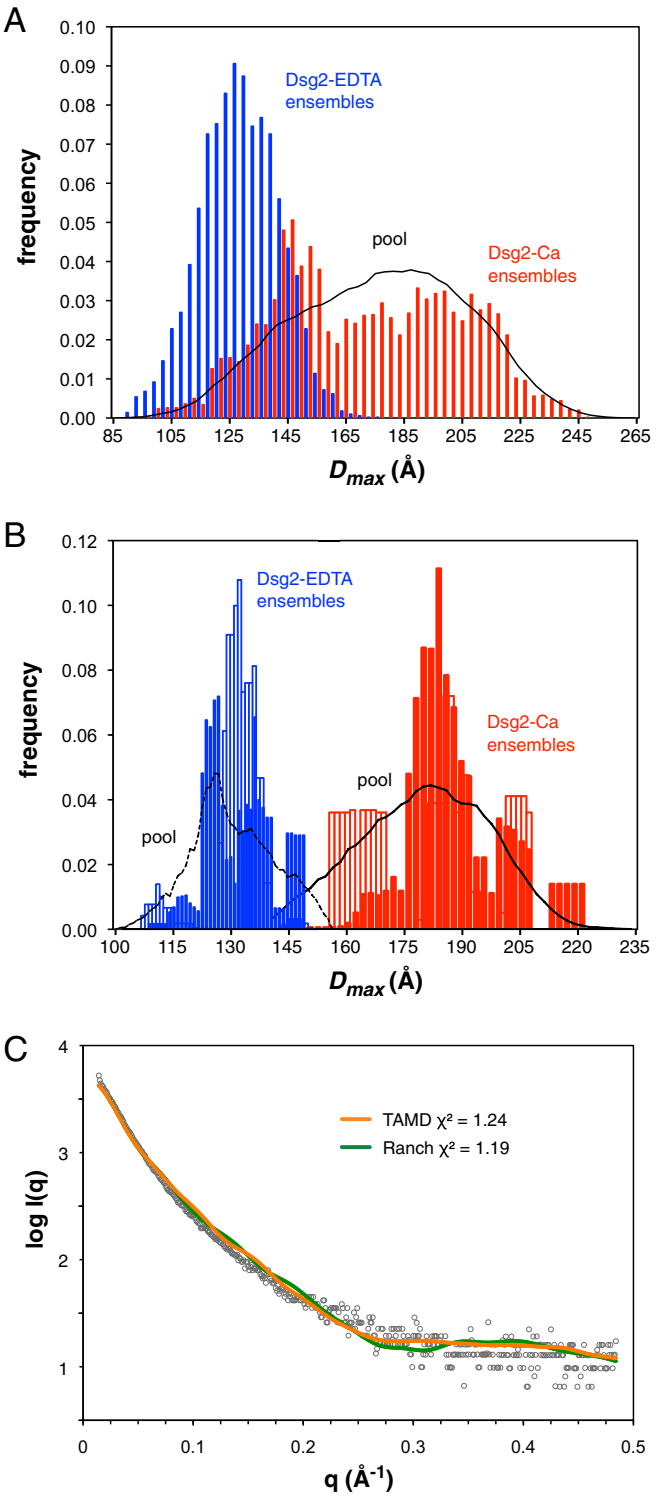


Fig. S4. Analysis of the EOM ensembles before and after filtering. (*A*) D_{max} distribution of the random pool and selected ensembles for Dsg2-EDTA (blue) and Dsg2-Ca (red). (*B*) D_{max} distribution after filtering the Ranch models (open bars) and TAMD (solid bars). (*C*) Fitting of the Mes ensemble from the filtered subsets Ranch (green) and CNS-TAMD (orange).

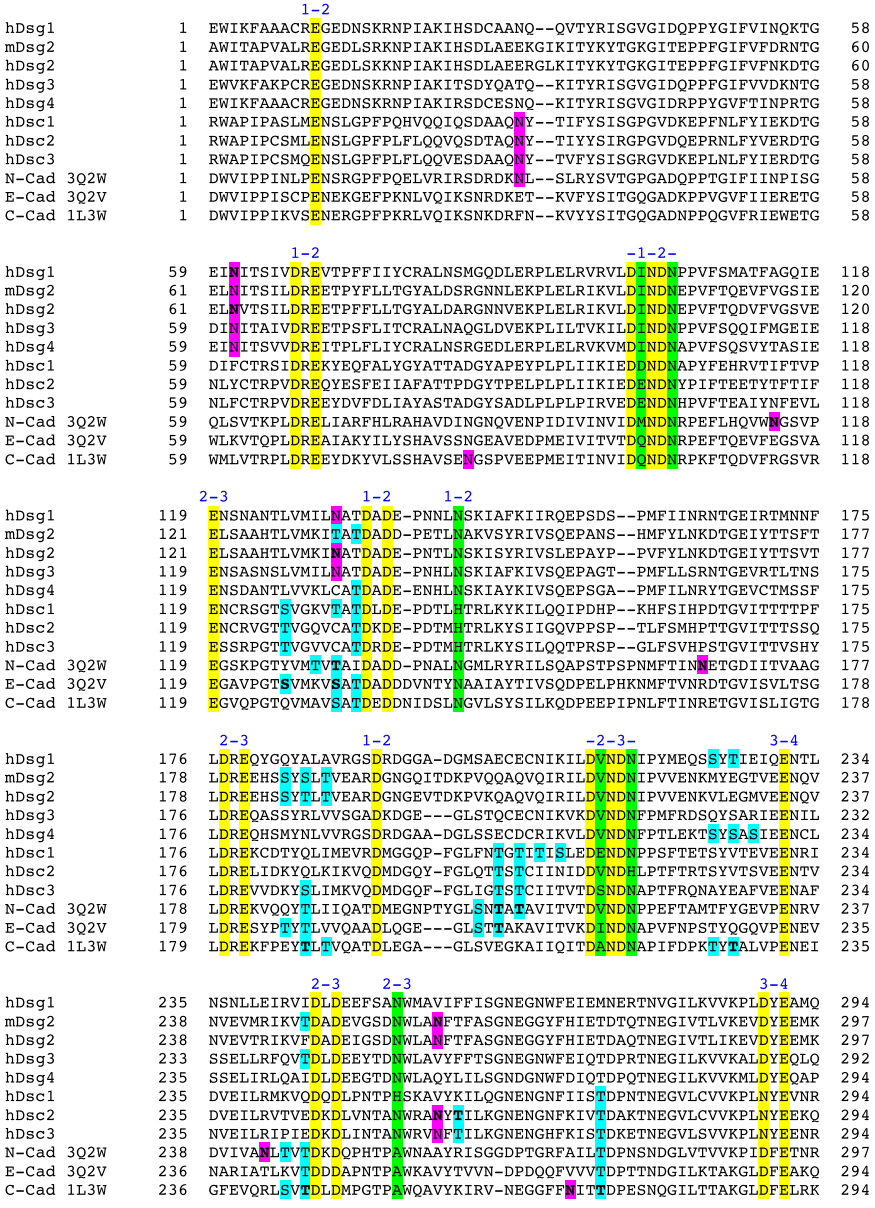


Fig. S5. (Continued)

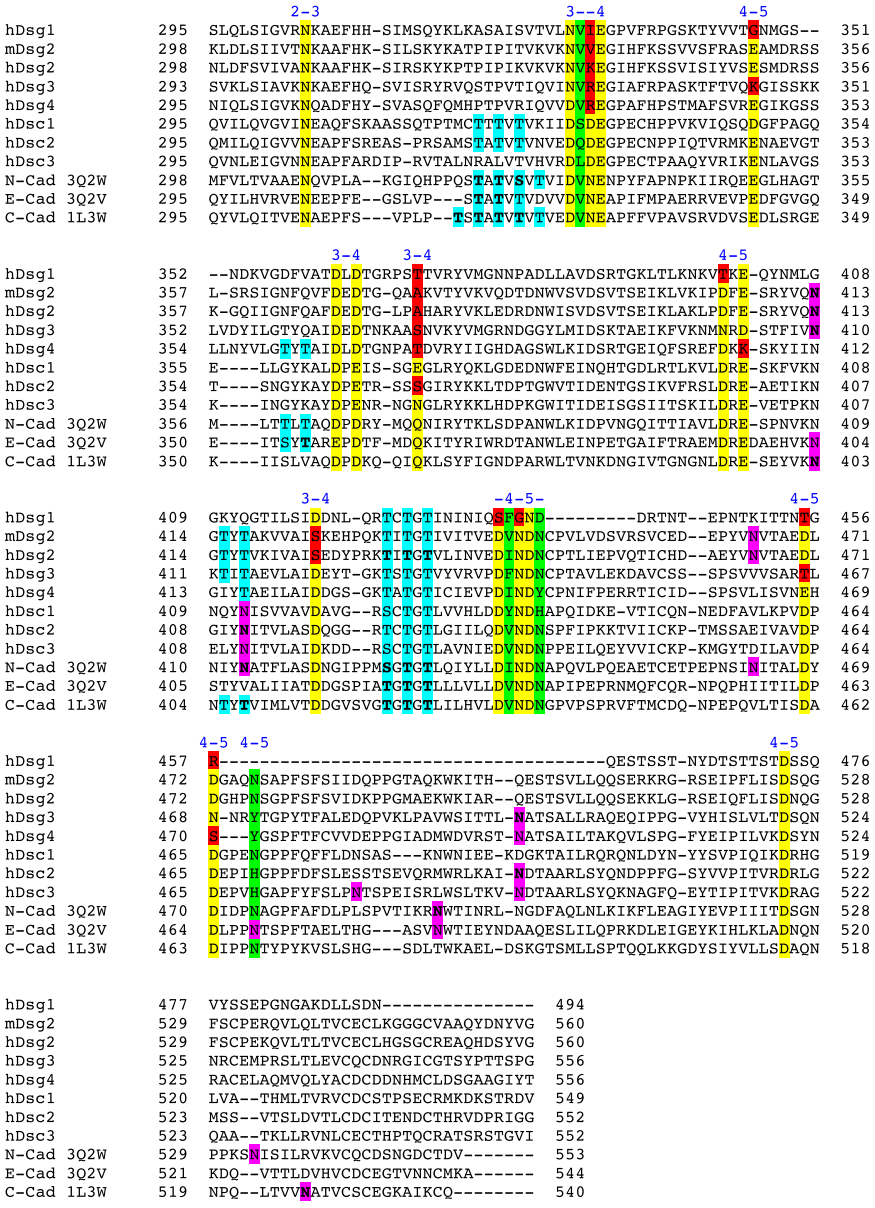


Fig. S5. Sequence alignments of desmosomal cadherins, including four human Dsgs (hDsg1–hDsg4), mouse Dsg 2 (mDsg2), three human Dscs (hDsc1–hDsc3), and type 1 N-, E-, and C-cadherins with known 3D full-ectodomain structure (indicated by their PDB ID codes). Calcium-binding residues are highlighted in yellow (side chain coordination) and green (main chain coordination). Numbering on top (blue) indicates the corresponding interdomain region (e.g., 1–2 is EC1–EC2). Residues with nonconserved side chain coordination are indicated in red. Predicted sites of *N*-linked and O-mannosyl glycosylation are highlighted in magenta and cyan, respectively; residues in bold correspond to glycosylation sites confirmed by mass spectrometry (7, 20) or X-ray crystallography (21, 22). Alignment made with ClustalX (23) and annotated manually.

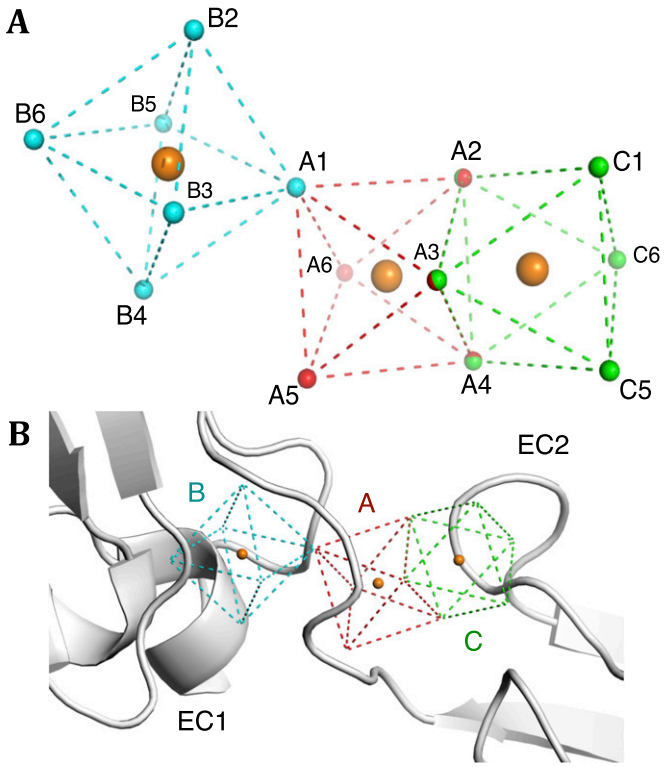


Fig. S6. Topology of the interdomain calcium binding sites, in type 1 cadherin extracellular domains, based on the structural information of the PDB structures 1L3W, 3Q2V, and 3Q2W (21, 22). (*A*) Each interdomain region contains three calcium-binding sites, A, B, and C, where each calcium ion is surrounded by six ligands in octahedral coordination. Ligand positions A1–A4 are shared by two calcium sites, and therefore need to be occupied by bridging side chains (Glu, Asp). No protein ligands are observed in positions C1 and C5, which presumably are occupied by water molecules. (*B*) Relative position of the three calcium-biding sites with respect to two neighboring domains, illustrated for the EC1–EC2 interdomain region.

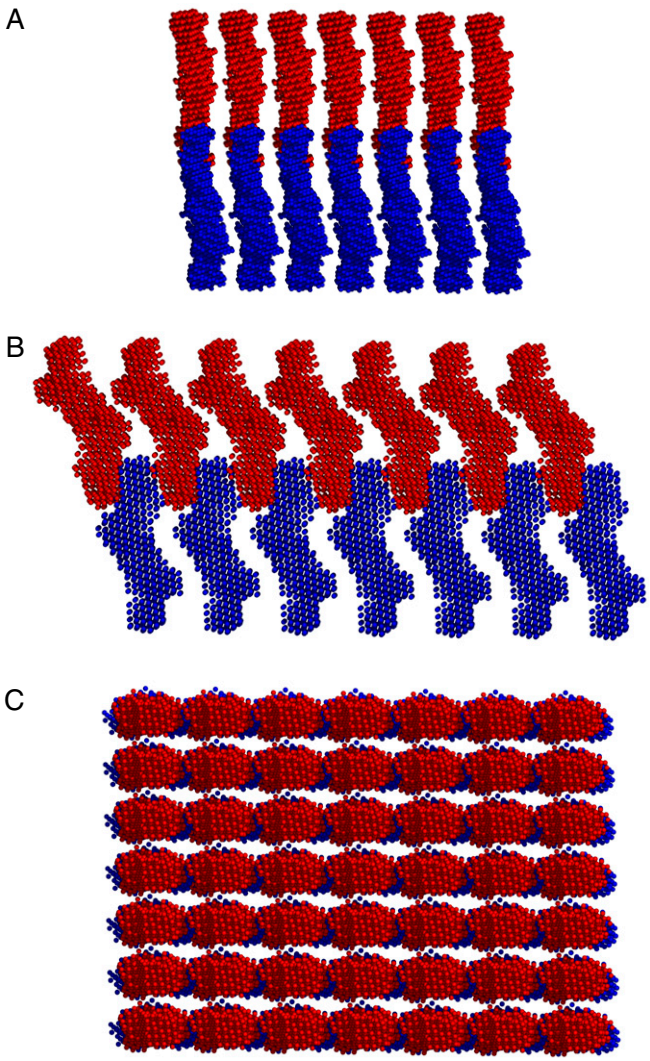
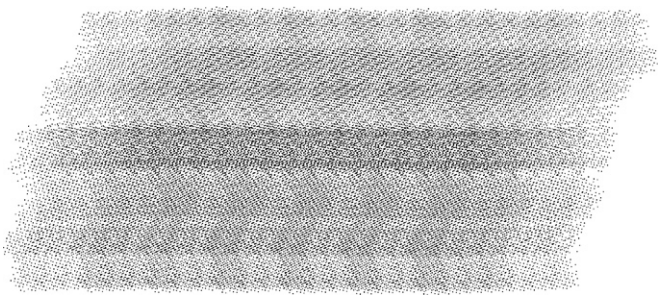


Fig. S7. Idealized array of Dsg-Ca ectomains. Front (*A*), side (*B*), and top (*C*) views of the array, generated with the DAMMIN models from the fitting into the ET maps.

Table S1. Calcium coordination in classical cadherin structures and equivalent residues in mDsg2 after multiple sequence alignment

	[———B site———]				[——— A site———]			[———C site———]				
Structures	B6	B5	B4	B3	B2	A1	A6	A5	A4	A3	A2	C6
EC1–EC2												
E-Cad	D195	ψ143	N102	ψ104	D134	D136	D100	ψ101	E11	D103	E69	D67
N-Cad	D194	ψ143	N102	ψ104	D134	D136	D100	ψ101	E11	D103	E69	D67
C-Cad	D195	ψ143	N102	ψ104	D134	D136	D100	ψ101	E11	D103	E69	D67
mDsg2	D194	ψ144	N104	ψ106	D136	D138	D102	ψ103	E11	D105	E71	D69
EC2–EC3												
E-Cad	N304	ψ254	N215	ψ217	D246	D248	D213	ψ214	E119	D216	E182	D180
N-Cad	N307	ψ256	N217	ψ219	D248	D250	D215	ψ216	E119	D218	E181	D179
C-Cad	N304	ψ254	N215	ψ217	D246	D248	D213	ψ214	E119	D216	E182	D180
mDsg2	N307	ψ256	N217	ψ219	D248	D250	D215	ψ216	E121	D218	E181	D179
EC3–EC4												
E-Cad	D415	Q365	N327	ψ328	E358	D360	D325	ψ326	E232	E328	E291	D289
N-Cad	D420	Q371	N333	ψ334	D364	D366	D331	ψ332	E234	E334	E294	D292
C-Cad	D414	Q365	N327	ψ328	D358	D360	D325	ψ326	E232	E328	E291	D289
mDsg2	**S424**	**A375**	**V334**	ψ335	D368	D370	N332	ψ333	E234	E335	E294	D292
EC4–EC5												
E-Cad	D517	ψ468	N435	ψ437	D462	D464	D433	ψ434	E343	D436	E397	D395
N-Cad	D525	ψ474	N440	ψ442	D468	D470	D438	ψ439	E349	D441	E403	D401
C-Cad	D515	ψ467	N434	ψ436	D461	D463	D432	ψ433	E343	D435	E397	D395
mDsg2	(D525)	(ψ476)	N444	ψ446	(D470)	(D472)	D442	ψ443	E350	D445	E407	D405

Site nomenclature as in Fig. S6. Main chain carbonyl binding is indicated with the ψ symbol. Residues in bold type are not conserved. Underlined residues are potentially conserved but appear difficult to model in a reasonable calcium-binding coordination. The dashed underlined residue can be modeled in binding coordination for only one of the two sites it is supposed to bridge. The C1 and C5 sites in Fig. S6 are water molecules completing the coordination of the third calcium ion.



Movie S1. Rotation around the vertical axis of the array of mDsg2 ectodomains, where the midline can be visualized.

Movie S1