Structural basis of the strong cell-cell junction formed by cadherin-23

Gayathri S. Singaraju1, Amin Sagar1, Anuj Kumar2, Jesse S. Samuel1, Jagadish P. Hazra1, Malay K. Sannigrahi1, Ragothaman M. Yennamalli3, Ashish4 and Sabyasachi Rakshit1,5

1 Department of Chemical Sciences, Indian Institute of Science Education and Research Mohali, Punjab, India
2 Department of Physical Sciences, Indian Institute of Science Education and Research Mohali, Punjab, India
3 Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, India
4 Institute of Microbial Technology (CSIR), Chandigarh, India
5 Centre for Protein Science Design and Engineering, Indian Institute of Science Education and Research Mohali, Punjab, India

Keywords
atypical cadherins; cadherin-23; cell–cell adhesion; single-molecule Forster resonance energy transfer; small-angle X-ray scattering

Correspondence
S. Rakshit, Department of Chemical Sciences, Indian Institute of Science Education and Research Mohali, Punjab, India
Tel: +91-7937476759
E-mail: srakshit@iisermohali.ac.in

(Received 21 February 2019, revised 13 August 2019, accepted 13 November 2019)
doi:10.1111/febs.15141

Introduction

Cell–cell adhesion by classical cadherins, a subfamily of cadherin class of proteins, is well-studied [1]. Non-classical cadherins that comprise more than 80% of cadherins also actively participate in cell–cell adhesion [2,3]. While the physiological significance of nonclassical cadherin-mediated cell–cell junction is well recognized, little is explored on their molecular structures. Cadherin-23 (Cdh23) is one of the giant nonclassical cadherins that forms strong intercellular junctions in nearly 90% of healthy epithelial tissues including the brain, lymph node, kidney, gastrointestinal tract, testis, and skin (The Human Proteome Atlas) [4–6]. The strong junction that is mediated by the homophilic trans-interactions of Cdh23 serves as metastasis suppressor for solid cancers including sarcoma, adrenal-cortical carcinoma, and cervical cancer (The Cancer Genome Atlas, TCGA) [6,7]. Interestingly, Cdh23 is also known for its strong binding with protocadherin-15 (Pcdh15) at the tips of stereocilia in neuroepithelial cells where the complex serves as gating spring under...
sound stimuli [5,6]. While the structural detail of the heterophilic complex with Cdh23 is understood [8], the molecular details of the homophilic complex are not identified yet. Here, we aim to understand the molecular mechanism of the homophilic interactions of Cdh23 that mediate a robust cell–cell adhesion.

Cdh23 comprises a cytosolic domain, a single-pass transmembrane region followed by 27 extracellular (EC) domains, unlike just 5 EC domains in classical cadherins [9–11]. However, similar to classical cadherins, Cdh23 causes cells to adhere using two types of homophilic interactions, trans and cis [4,12]. Electron tomography revealed a unique pattern for the cis-homodimer of Cdh23: a pair of Cdh23 molecules aligned in the same orientation and intertwined to form a helical complex through interactions between all EC domains except few terminal domains [5]. The domains at the N termini that are exposed outward are available for trans-interactions (Fig. 1A). Using single-molecule Forster resonance energy transfer (smFRET) and live-cell aggregation assays, here we first identified the number of domains that participate in the trans-dimerization. In relevance, the engagement of the two terminal domains in a heterophilic trans-interaction with protocadherin-15 (Pcdh15) was already reported in tip links in the inner ear [8].

Classical cadherins undergo homophilic trans-interactions using the outermost terminal domain, EC1 [13,14]. They first form an X-dimer, a kinetically driven interaction and then converted to a thermodynamically stable strand-swap dimer (S-dimer) via an intermediate with the overlap of the linker region between two terminal domains [15,16]. Among non-classical cadherins, desmosomal cadherins form an S-dimer [17], whereas T-cadherin and R-cadherin form only X-dimer [18]. Cdh23, however, lacks the sequence determinants for either S- or X-dimerization. Moreover, the EC1 domain of Cdh23 has several unique
features: a 5 residue long $3_{10}$-helix just prior to the A* $\beta$-strand, an $\alpha$-helical loop connecting two $\beta$-strands, and most strikingly, an additional Ca$^{2+}$-binding site toward the N terminus [19]. Together, these features indicate a unique interface for Cdh23-mediated homophilic trans-interactions. It was therefore imperative to decipher the molecular details of the Cdh23-mediated homophilic trans-interactions.

Using small-angle X-ray scattering (SAXS) in combination with in silico measurements including docking and molecular dynamics, we identified the molecular structure of the trans-homodimer of Cdh23. We verified the binding interface with a single point mutation that impaired the dimer. Finally, using analytical methods including ultracentrifugation (AUC) and single-molecule force spectroscopy (SMFS) with an atomic force microscope (AFM), we estimated the thermodynamic and kinetic parameters of the homophilic complex.

**Results**

**The homodimer of Cdh23 is a trans-dimer interacting via the N termini of the two outermost domains (EC1-2)**

Two outermost domains of Cdh23 are known to form the trans-heteromeric complex with Pcdh15 in tip links. In order to determine the number of EC domains required for the trans-homodimerization, we expressed Cdh23 with varying lengths of EC domains: first domain alone (EC1), first two domains (EC1-2), and first three domains (EC1-3). All the constructs were expressed in *E. coli* BL21 RIPL following a reported protocol (Materials and methods) [19] and purified in two steps by Ni$^{2+}$-NTA-based affinity followed by size exclusion chromatography (SEC) (Fig. 1B). We ran SEC for all the three constructs at a high concentration (100 µM) and observed two distributions in elutions, P1 and P2, for EC1-2 and EC1-3 and EC1 eluted as a monomer (Fig. 1B). We subsequently ran SDS/PAGE for all the eluted fractions and observed a single band at 26 kDa (Fig. 1C) for Cdh23 EC1-2 and 38 kDa for Cdh23 EC1-3, suggesting that the P2 and P1 corresponded to the monomer and a higher-order association, respectively. To determine the molecular weights of the elutions, we developed a calibration curve for SEC with standard proteins of varying molecular weights under the same conditions (Fig. 1D). From the standard curve, we then estimated the apparent molecular weights of proteins eluted at P2 and P1 fractions. The apparent molecular weights corresponded to 27 and 51 kDa for Cdh23 EC1-2 and 45 and 76 kDa for Cdh23 EC1-3, respectively. These estimated molecular weights corroborated with the theoretical monomer and dimer molecular weights, 26 and 52 kDa for Cdh23 EC1-2 and 39 and 79 kDa for Cdh23 EC1-3, respectively, confirming a dimer in the higher-order association. The negligible variations in the apparent molecular weights, as opposed to the theoretical values, may arise from the differences in shapes of Cdh23 than the standard globular proteins. Further, we noticed significantly higher intensity for P1 than P2 for Cdh23 EC1-2 and reversed for Cdh23 EC1-3, though the loading concentration of proteins was the same for both the SEC runs (Fig. 1B). Thus, the higher intensity of P1 than P2 for Cdh23 EC1-2 is indicative of the highest binding affinity for Cdh23 EC1-2 toward homodimerization than the other constructs.

Next, we performed smFRET to decipher the orientations and the extent of overlap of the constituent monomers in the dimer. We used Cdh23 EC1-2 and Cdh23 EC1-3 for the experiment and not Cdh23 EC1, as Cdh23 EC1 showed the weakest affinity toward dimerization in the SEC. Since we performed smFRET on a glass coverslip using a total internal reflection fluorescence microscopy (TIRFM), we covalently anchored the C terminus of the proteins, Cdh23 EC1-2, and Cdh23 EC1-3 individually, to the coverslips using sortagging chemistry as reported [20,21]. All protein constructs were recombinantly modified with the sortase recognition sequence (-LPETGSS) at the C terminus. Prior to the surface attachment, proteins were modified with a donor (D) dye, Cyanine3 (Cy3, $\lambda_{em} = 545$ nm; $\lambda_{ex} = 560$ nm), at the N terminus. For N-terminal modification, we recombinantly mutated the valine at position 3 to cysteine (V3C) and attached Cy3 using the thiol–maleimide Michael addition reaction. Thus, irrespective of the Cdh23 constructs, the surface-anchored proteins always had donor dyes at the N terminus. Next, the protein modified surface was incubated with the second batch of proteins for 30 min for the homodimers to form, followed by vigorous washing to remove nonspecific attachments. The protein in solution was tagged with acceptor (A) dye, Cyanine5 (Cy5, $\lambda_{ex} = 645$ nm; $\lambda_{em} = 660$ nm), either at the C terminus or N terminus. The N-terminal modification was done using V3C protein constructs, whereas the C-terminal modification was done using sortagging followed by the thiol–maleimide Michael addition (Materials and methods). For all the constructs, the final dye to protein ratio was maintained at 1 : 1.

smFRET was measured for two different combinations of each protein construct, N$_D$N$_A$ and N$_D$C$_A$. 

---

**G. S. Singaraju et al.**

---

---
ND2NA is N-terminal donor (D) with N-terminal acceptor (A) dye, and ND2CA is N-terminal donor (D) with C-terminal acceptor (A) dye. Altogether, we have four such combinations, ND2NA and ND2CA for Cdh23 EC1-3 homodimers (Fig. 2A(i)) and Cdh23 EC1-2 (Fig. 2A(ii)) homodimers. We excited the donor molecules with a 532 nm laser and subsequently monitored emissions in donor and acceptor channels for 400 s using an EMCCD (Materials and methods). We used rms software (The Birkedal lab, Aarhus Universitet) for the analysis of FRET traces at each colocalized spot of single donor and single acceptor (Materials and methods) [22]. The FRET efficiency ($E_{\text{FRET}}$) was estimated from the intensity ratios between acceptor and donor and plotted as distributions (Fig. 2B, left panel). The unimodal distribution of $E_{\text{FRET}}$ for all the combinations indicates a single conformational population of the dimer (Fig. 2B, right panel). We used the peak maxima of the distributions as most probable efficiency ($E_{\text{FRET}}^{\text{mp}}$) and estimated the donor–acceptor separations for each $E_{\text{FRET}}^{\text{mp}}$.

The maximum $E_{\text{FRET}}^{\text{mp}}$ was measured for the ND2CA combination of Cdh23 EC1-2 homodimers with an $E_{\text{FRET}}^{\text{mp}}$ of 0.60 ± 0.007, indicating the closest proximity between the N terminus and C terminus of Cdh23 EC1-2 constructs (Fig. 2C). Interestingly for all other combinations, we measured lower $E_{\text{FRET}}^{\text{mp}}$ values with negligible differences. The closer association of the ND2CA than the ND2NA termini in Cdh23 EC1-2 confirms the trans-conformation of the dimer (Fig. 2C). The lower $E_{\text{FRET}}^{\text{mp}}$ for the ND2CA combination of Cdh23 EC1-3 than Cdh23 EC1-2 indicates that the overlap extends only up to EC1-2 domains. Finally, no significant differences in the $E_{\text{FRET}}^{\text{mp}}$ distributions of Cdh23 EC1-3 for ND2CA and ND2NA combinations support an extended overlap of the EC1-2 domains in the trans-homodimer as depicted in the models (Fig. 2A, Table S1). The comparable $E_{\text{FRET}}^{\text{mp}}$ of ND2NA for Cdh23 EC1-2 with Cdh23 EC1-3 further supports the model of extended overlap between EC1-2 domains in the trans-homodimer (Fig. 2A(i) and Table S1). Thus, we conclude from our smFRET measurements and SEC results that Cdh23 forms trans-homodimer with the two outermost domains (EC1-2) alone. In all our following experiments, we used Cdh23 EC1-2 only, unless mentioned.

**Cdh23 EC1-2 can mediate cell–cell adhesion**

A549 cells endogenously express Cdh23 (full-length, variant-1) [6]. To check the functional significance of Cdh23 EC1-2 concerning the full-length construct, we measured the *in vitro* binding of A549 live cells on Cdh23 EC1-2-modified surfaces. For the cell adhesion experiment, C terminus of the Cdh23 EC1-2 proteins was covalently attached on the coverslips using sortaging and then incubated with ~ 10^7 numbers of A549 live cells for 2 h, followed by repeated gentle washes with a buffer containing 2 mM of Ca^{2+}-ions. We subsequently imaged the coverslips under bright field and observed a large number of cells remained adherent to the surface (Fig. 2D, left panel). Since the cadherin-mediated interactions are Ca^{2+}-dependent, as a control for the specificity of the cell surface interactions, we monitored the number of cells adhered to surfaces after washing with Ca^{2+}-free EGTA buffer (Fig. 2D, right panel). We noticed ~ 86% reduction in the number of cells adhered to the surface (Fig. 2D, Inset) as expected. We further quantified the number of live cells adhered to the surface by colorimetric MTT assay and measured 55.8% fewer cells for EGTA-washed coverslips (Fig. 2E). These studies corroborate with the smFRET results, supporting that EC1-2 domains of Cdh23 are enough to mediate the Ca^{2+}-dependent cell–cell adhesion. We extended the demonstration from the aggregation of live cells (HEK-293) transfected transiently with Cdh23 EC1-2 [23]. Untransfected HEK-293 cells were used as a control. For both, cells were pretreated with Cdh23-specific siRNA to silence the endogenous expression of Cdh23. We observed that cells transfected with Cdh23 EC1-2 formed aggregates within 120 min (Fig. 2F(i)), while the untransfected HEK-293 cells did not show any aggregates (Fig. 2F(ii)).

**SAXS envelope portrays a compact extended-handshake conformation for the dimer**

To reveal the shape of the Cdh23 EC1-2 trans-homodimer in solution, we performed SAXS measurements with proteins at 10 mg·mL^{-1} (Fig. 3A), for a q-range of 0.01 to 0.45 Å^{-1} using SAXSpace instrument (Anton Paar GmbH, Graz, Austria) (Table S2). The monodispersity of the sample was verified from the linearity of the Guinier plot (Fig. 3A, inset). The folding of the proteins was confirmed from the strong agreement between the normalized Kratky plots ($\langle I(q) \rangle$ vs $q^2 \langle I(q) \rangle$) of the experimental data and the theoretical SAXS profiles of the crystal structures/models (Fig. 3B, inset). It is pertinent to mention here that the peaks of the normalized Kratky plots deviate from the value of 1.73. This can be attributed to the rod-like shape of the protein, as the theoretical SAXS profile of the docked model also showed the same profile (Fig. 3B, inset) [24]. The molecular weight was estimated as 54 kDa from the volume of correlation ($V_c$)
Trans-homodimer structure of cadherin-23

G. S. Singaraju et al.

A i) Cdh23 EC1-3

ii) Cdh23 EC1-2

C

d_{NC} < d_{NN}

E

WT-Ca^{2+}

% Cell adhered to surface

0 20 40 60 80 100

EGTA

F i

Cdh23 EC1-2

0 min 120 min

ii

Untransfected

0 min 120 min

Ca^{2+} - Buffer EGTA-Buffer
Fig. 2. Homodimerization of cadherin-23 in the trans-conformation. (A) The probable orientations of the proteins, (i) Cdh23 EC1-3 and (ii) Cdh23 EC1-2, in the dimeric forms are represented schematically. The conformations are named as N2D2 or N2C2 according to their dye-labeled terminals. (B) Left panel represents the time traces of the fluorescence intensities for the donor (D, green) and acceptor (A, red) dyes for four different complexes, two with Cdh23 EC1-2 and the rest two with Cdh23 EC1-3. For both sets of homodimers, the complexes are marked differently as N2D2 and N2C2 as per the labeling terminals with the donor (D) and acceptor (A) dyes. For each set, we analyzed 25 spots that showed FRET (N = 25). All the FRET experiments were repeated twice. The corresponding FRET efficiency distributions are plotted in the right-side panel. (C) Schematic depiction of the distances between the interacting proteins, $d_{AC}$ and $d_{AP}$. (D) Left, Pull-down of live A549 cells expressing Cdh23 (variant-1) endogenously, by the surface-anchored Cdh23 EC1-2 proteins in the presence of Ca$^{2+}$ buffer. Right, The cells detached upon chelating out Ca$^{2+}$ ions using EGTA (scale bar: 60 μm). The inset shows the histogram of the number of cells adhered to the surfaces in the presence (gray) and absence (black) of Ca$^{2+}$. The number of cells was estimated from the grayscale intensity of the images using ImageJ software. (E) The scattered plot represents the percentage of adhered A549 cells to the surfaces, estimated from the MTT assay. The percentages are measured from five different surfaces for each sample, and each experiment was repeated 3 times, which makes $N = 15$. The error bars represent the SEM with $N = 15$. (F) (i) Time-dependent cell aggregation of HEK-293 cells transiently transfected with Cdh23 EC1-2 is captured using bright-field imaging along with the control of (ii) untransfected HEK-293 cells (scale bar: 75 μm). For both experiments, the cells were transfected with Cdh23 siRNA to stop the expression of endogenous Cdh23. The images were captured over five different surfaces, and the experiments were repeated 3 times, which makes $N = 15$.

[25] and 60 kDa by lysozyme as standard. From the indirect Fourier transformation of the SAXS profile using GNOM [26], we generated the pair distribution function ($P(r)$), which provided the frequency of the interatomic vectors inside the predominant shape of the proteins in real space ($r$). $P(r)$ curves estimated for dimer showed a maximum linear dimension ($D_{max}$) and $R_g$ values of 11.6 nm and 3.2 ± 0.2 nm, respectively (Fig. 3B). The $D_{max}$ further supports the extended-handshake conformation for the dimer as predicted from smFRET. The peak position and shoulder profile of the computed $P(r)$ supported the multidomain shape of Cdh23 connected by a nonflexible linker.

To obtain a three-dimensional shape of the dimer, we generated ten independent dummy residue models using DAMMIF [27], averaged using DAMAVER, refined using DAMMIND [28], and compared with each other by calculating the normalized spatial discrepancy (NSD). NSD is a measure of the similarity in the shapes of the models with a value of < 1 indicating that the models are relatively similar and values above 1 indicating that the models are systematically different from each other. The mean NSD between the 10 models was 0.602 with a standard deviation of 0.027, indicating that all the models were very similar to each other. We averaged all ten models and obtained an envelope with dimensions of 14(L) × 5.76(W) × 4.1 (H) nm (Fig. 3C). We repeated the modeling protocol several times and validated the protocol as robust and reproducible.

Homodimerization of Cdh23 EC1-2 is not mediated by tryptophan

Next was to identify a dimer structure that could fit the SAXS envelope. We first considered the W-conformation of Cdh23 dimer proposed previously [19]. It was predicted that Cdh23 EC1-2 might form a trans-homodimer through π-stacking of the indole ring of the sole tryptophan at 66th position (W66). The driving force behind such W-mediated interactions is the switch of the W-environment from hydrophilic to more hydrophobic. Since W is known to feature a solvatochromic shift in fluorescence, we designed experiments to probe the W-emission of the monomer and dimer and decipher its role in dimerization. Accordingly, we monitored the steady-state emission of W66 ($λ_{ex} = 295$ nm). As the protein concentration increased from a monomer concentration to beyond the $K_D$ (~18 μM) of the dimer (Fig. 3E), we did not observe any shift in W-emission, excluding its possible involvement in dimerization. Further, we examined the mobility of W66 in the monomer and dimer using time-resolved fluorescence anisotropy. Trp in a protein can have two rotational components, a fast rotation around its axis and a relatively slower rotation along with the protein. Dimerization via the π-stacking of the W residues is expected to constrain the free local rotation around its axis and therefore delay the anisotropy decay of the fast component. We probed this anisotropy decay of W66 at two concentrations of proteins ranging from monomer populated solution to dimer dominated solution (Fig. 3F) and observed no change in the decay rate of the faster components. This indicates a lack of constraints on the local mobility of W66 and thus confirming that W66 did not play a role in dimerization. However, we observed a significant increase in the decay for the global rotation of W66 upon dilution indicating a decrease in overall size. The decay at low concentration matches with the global rotation of Cdh23 in the presence of EGTA, which is expected to be in the monomeric form (Fig. 3F and Table S3).
Once we confirmed the W-conformation cannot be the dimer, we focused on searching for new structures by docking Cdh23 EC1-2 (PDB ID: 2WHV) using PatchDock. Of the first ten models based on docking scoring, seven models showed docking through the N terminus of the protein and the others through the C terminus. We considered these seven models and superimposed them with the SAXS envelope using SUPCOMB, which showed comparable NSD values varying from 0.7 to 0.8 for all the structures. However, rank 1 appeared as the best fit model from the z-tests based on $R_S$ and FRET-based end-to-end distances ($P = 0.20$) (Fig. 3D and Table S4).

We next performed molecular dynamics (MD) simulations of the rank 1 for 100 ns using GROMACS and delineated residue-wise interactions (Materials and methods, Table S5). Stability of the dimer was achieved within 2 ns of the simulations to an average

---

**Fig. 3.** Cadherin-23 dimer forms an extended-handshake conformation. (A) SAXS profile of the Cdh23 EC1-2 dimer is plotted as the scattering intensity $I(q)$ vs. the scattering vector $q$ on a log–log scale. The solid black line represents the theoretical SAXS profile computed from the proposed model of the dimer. The average scattering intensity is collected by exposing the sample to radiation for 20 min in 3 cycles. The acquisition was performed for 5 different batches of proteins, which makes $N = 15$. The linear nature of the Guinier plot ($\ln(I(q))$ vs. $q^2$) for the low q region is marked (inset, bottom, left). (B) The $p(r)$ distribution for the dimer (blue) showing the maximum linear dimensions ($D_{max}$) to be 11.6 nm. The normalized Kratky plot ($I(q)/I(0)^*q^2R_g^2$ of the experimental SAXS profile superimposed with that of the theoretical SAXS profile of the proposed dimer is shown in the inset. (C) The average ab initio model obtained from DAMMIN is shown for the dimer. (D) The statistically verified and rank-1 PatchDock structure (PDB: 2WHV) of the dimer (blue) was fitted to the SAXS envelope (gray) for the dimer and shown with two orthogonal perspectives. (E) Steady-state fluorescence anisotropy decays of W66 in Cdh23 EC1-2 were recorded at two concentrations of protein, 11 µM (red) to 44 µM (blue), in Ca$^{2+}$-buffer. Increasing protein concentrations increase the dimer population in the solution. The black dots represent the fluorescence anisotropy decays of 44 µM Cdh23 EC1-2 in 1 mM EGTA where the protein is expected to be in monomeric form exclusively. We confirmed the reproducibility of the result by repeating all the experiments 3 times for each concentration, both 11 and 44 µM, using different batches of protein.
root mean square deviation (RMSD) of 0.34 nm (Fig. 4A,B). To identify the residues that are involved in dimerization, we compared the root mean square fluctuations (RMSF) of all residues between the monomer and dimer anticipating that residues buried within interacting surfaces may fluctuate less (Fig. 4C). The residues other than glycine that showed significant fluctuation differences are highlighted (Fig. 4C, inset). The time trace analysis of the MD results also indicated these residues responsible for the interactions.

Fig. 4. Survey of the H bonds that are crucial for dimerization. (A) Superimposed snapshots of the Cdh23 EC1-2 dimer at 10-ns intervals of the entire trajectory of the MD run are shown as a visual guide to the RMSD plot. (B) RMSD for the rank-1-docked Cdh23 EC1-2 dimer is shown for 3 independent MD runs of 100 ns. (C) RMSF is shown to compare the average fluctuations ($N = 3$) of all residues in chain A (magenta) and chain B (blue) of the dimer with the monomer (wine red). The zoomed plots in the inset highlight the RMSF of the interacting residues (left: K76 (purple), S77 (green), E78 (red) from interface 1 and right: N112 (blue) and P114 (yellow) from interface). (D) The plot shows the survival probabilities of the H bonds over time between different residues. From bottom to top, the interaction residues are divided into interface 1 and interface 2. The residues in pink are from chain A, and the blue residues are from chain B of the dimer. (E–J) The images show the relative position of the residues involved in H bonding in enlarged diagrams between two chains, Chain A (magenta) and Chain B (blue): (E) H bonds between N112–Q2; (F) Q2–S116; (G) R5–P128; (H) E78–S77; (I) K76–E78; and (J) Q99–S77. The black dots represent the H bonds.
We divided these residues into two main interacting interfaces based on their positions in the EC domains (Fig. 4E–J). Interface one is dominated by electrostatic interactions that are mediated by the antiparallel overlap of strand F of the EC1 repeats. The residues involved in the interactions are K76, S77, E78, N97, and Q99, which are conserved in human, mouse, rat, and zebrafish (Fig. 5). Interestingly, S77 was found mutated into L (leucine) in patients who have cutaneous cancer. Some of these residues are also instrumental in heterodimerization with Pcdh15 at tip links. The other interface, which is amphiphilic, is formed by anchoring of the elongated N-terminal strand of EC1 of one monomer to the β-strand (G-strand) of EC2 of the other protein. None of these residues are occluded by glycosylation, indicating that the interfaces are physiologically relevant (Fig. 6A). In vitro cell-binding assays also showed the arrest of A549 cell lines on surfaces coated with Cdh23 EC1-2 wild-type (WT) with no post-translational modifications. To understand the stability of the interface, we introduced a single point mutation at E78 to a positively charged residue, K (E78K). Proper folding of the mutant was verified using SEC and circular dichroism (Fig. 6B,C). To check the formation of the homodimer with mutants on a surface, we performed a live-cell binding assay using A549 cells incubated on a glass-surface 96-well plate premodified with Cdh23 EC1-2 (E78K) as described before (Materials and methods). We observed a significant decrease (78% drop) in the number of cells adhered to Cdh23 EC1-2 (E78K) modified surfaces than WT (Fig. 6D). Quantitative estimation using MTT assay also measured a 35.8% decrease in the total number of viable cells bound to surfaces than Cdh23 EC1-2 (WT)-coated surfaces, suggesting the lack of homophilic interactions for the mutant (Fig. 2E). Similarly, we observed the complete disruption of the dimer structure for the E78K mutant in solution even at a concentration of 10 mg·mL⁻¹ (440 μm). The \( R_g \) of E78K obtained from SAXS is 3.0 ± 0.2 nm, corresponded to a monomer (Fig. 7A). The indirect Fourier transform gave a \( D_{max} \) of 10 nm (Fig. 7B). Further, we were able to fit the SAXS envelope (Fig. 7C) obtained for E78K to the crystal structure of the monomer (Fig. 7D). These values correlated well with the \( R_g \) and \( D_{max} \) of the monomeric WT protein (2.9 and 10.0 nm, respectively) as observed in the X-ray crystal structure (PDB ID: 2WHV). Also, the molecular weight was estimated to be 23 kDa based on \( V_c \) and 30 kDa based on the scattering intensity of lysozyme. The deviation from the expected mass while using lysozyme as standard might be arising from an error in concentration. More importantly, both the methods estimated the molecular weight of Cdh23 EC1-2 (E78K) to nearly half of the native, Cdh23 EC1-2 dimer. We were able to fit the SAXS envelopes (Fig. 7C) obtained for E78K to the crystal structure of the monomer and the docked structure by aligning their inertial axis using SUPCOMB (Fig. 7D).

**The trans-homodimer of Cdh23 EC1-2 has a high dissociation constant, however, is long-lived**

Once we deciphered the molecular structure of the homodimer of Cdh23, we were interested to measure the kinetic and thermodynamic parameters of the complex. We performed sedimentation velocity (SV) experiments with Cdh23 EC1-2 (WT) (Fig. 7E-I) using analytical ultracentrifugation (AUC) at two protein concentrations, 10 µM, and 36 µM (Fig. 8A–D), with a rotor speed of 142 000 g at 20 °C. The SV run at the higher concentration produced two distributions of sedimentation coefficient (C(s)). The peak that appeared at 2.1 s corresponds to the monomer, while the peak at 4.77 s with a lower frictional coefficient ratio (f/f₀) of 1.34 corresponds to the dimer (Fig. 7E(i)). Further, the dimer was confirmed from the estimation of the molecular weight (52 kDa) and \( R_h \) (4.1 ± 0.3 nm). The similar SV run for Cdh23 EC1-2 (E78K) (Figs 7E(ii) and 8E,F), however, featured only single distribution for a monomer even at a concentration of ~ 440 μM (Fig. 7E(ii)). The \( R_h \) estimated for Cdh23 EC1-2 (E78K) was 2.9 ± 0.1 nm (Fig. 7E(ii)).

We then performed sedimentation equilibrium (SE) using AUC with 40 µM of Cdh23 EC1-2 (WT) protein.
and 60 µM) at three different rotational speeds (46,000, 63,000, and 82,000 g) and estimated the $K_D$ of monomer–dimer equilibrium as 18 ± 4 µM from the global fitting of the equilibrium curves using free software SEDPHAT (Fig. 8I) [29]. The $K_D$ obtained for Cdh23 EC1-2 (WT) is significantly lower than E-cadherin (96.5 ± 10.6 µM) [30], however, comparable with N-cadherin (25.8 ± 1.5 µM) [30] and R-cadherin (13.7 ± 0.2 µM) [31]. Interestingly, we observed only one dominating peak with a $C(s)$ of 1.96 s and a $f/f_0$ of 1.51 corresponding to a monomer when we ran SV with 10 µM for Cdh23 EC1-2 (WT) (Fig. 8C,D), even though the dimer-to-monomer ratio at 10 µM of protein is nearly 1/4th for a homodimer with $K_D$ of 18 ± 4 µM. No detection of the dimer here could be an artifact from the detection limit of the instrument or could be due to slow binding on-rate of Cdh23 EC1-2 (WT) toward dimerization [32]. Moreover, it is well documented that the $K_D$ obtained from SV or SE varies widely for the same proteins and these variations are mainly attributed to the differences in internal pressure in AUC cell arising from different run-time [33–36].

Since the interface one is mediated by EC1 alone, we checked whether the interface one interacts independently to form the homodimer. As expected, SV with Cdh23 EC1 (WT) showed a population of both monomers and dimers indicating that EC1 alone can form a homodimer with a lower affinity ($K_D$ ~ 52 ± 7 µM, Fig. 8G,H,J) than EC1-2. Interestingly, we did not observe any dimer in SEC when injected 100 µM of Cdh23 EC1 which should contain 80% of the dimer. The nonappearance of the dimer peak in SEC could be due to in-column dilution of the protein before elution.

Finally, we performed SMFS with Cdh23 EC1-2 (WT) to estimate the strength of the interfaces as in cellulo studies measured strong adhesive properties of Cdh23 that suppresses tumor metastasis. For SMFS, we covalently attached the C terminus of the protein to AFM cantilevers (Si$_3$N$_4$) and glass coverslips using the sortagging protocol [37] (Fig. 9A, Materials and methods). A typical single-molecule unbinding force curve (unbinding event, black line) and a no-event (blue) curve that is obtained for SMFS
are shown in Fig. 9B. Overall, we observed 8–10% unbinding events, 97% of which featured single unbinding force curves that fit the freely jointed chain (FJC) model (Fig. 9B). These data corroborate well with the Poisson statistics used for single-molecule sorting. We plotted the maximum unbinding forces for seven different loading rates as histograms (Fig. 9C). The contour length ($L_c$) was estimated as an FJC fitting parameter for each of the individual unbinding curves, which distributed as a single Gaussian centered at 58.7 ± 1.0 nm irrespective of the pulling rate (Fig. 9B, inset). This finding is typical for the stretching of two PEG molecules of 5 kDa molecular weight in series. The unbinding force distributions for Cdh23 EC1-2 (WT), however, showed two well-separated binomial distributions (Fig. 9C). We hypothesized that the two distributions correspond to two different binding states: The low-force distribution is due to fast binding by either of the two interacting interfaces, and the high-force distribution is due to a complete...
handshake binding between Cdh23 EC1-2 (WT) partners. For verification, we repeated the force spectroscopy with Cdh23-EC1(WT) alone and measured the strength of interface one. As confirmation, the unbinding force distribution obtained for EC1 alone (red, Fig. 9C) perfectly overlaid with the low-force distribution of Cdh23 EC1-2. We also measured a higher probability of events for Cdh23-EC1 (WT) (~10%) than for EC1-2 (WT) (~6%). This may refer to a higher binding on-rate of Cdh23-EC1 (WT) than Cdh23 EC1-2 (WT).

We then plotted the most probable forces ($F_{mp}$) obtained from the peak maxima of the distributions with increasing loading rates. The loading rate for each velocity was estimated by considering the molecular tether and cantilever in series (Materials and methods) [38,39]. The data were fitted to the Bell–Evans model (Eq. 6) [40,41], and the intrinsic lifetime ($\tau_0$)
and the minimum distance from the bound state to the transition state ($x_b$) were determined (Fig. 9D,E). For Cdh23 EC1-2 (WT), we measured two $F_{mp}$ values for two force distributions at each loading rate. Our results indicate a very high $t_0$ of 1224.2 s for Cdh23 EC1-2 (WT) for the high-force distribution and $x_b$ of 0.31 nm. The $t_0$ and $x_b$ obtained for the low-force distributions are 1.86 s and 0.55 nm, respectively, which corresponded to the values obtained for Cdh23-EC1 (WT) alone with a 95% confidence interval, corroborating with the SV run. Next, we repeated the force spectroscopy experiments with Cdh23 EC1-2 (E78K) mutant and measured very low frequency of events similar to nonspecific measurements confirming again that Cdh23 EC1-2 (E78K) mutant does not interact with each other.
**Discussion**

Cdh23 mediates stronger cell–cell adhesion via homophilic trans-binding than classical E-cadherin [10]. The implication of such strong adhesion by Cdh23 in physiology is metastatic suppression [6]. It, therefore, became imperative to decipher the structural basis of the strong adhesion. From a barrage of biophysical and biochemical data, here we presented the first view of the homophilic trans-binding structure of Cdh23. The binding interface is mediated by the extended antiparallel overlap of the two terminal domains covering a surface area of 1796.3 Å² per monomer. Experiments including analytical SEC at 4 °C, smFRET at 20 °C, AUC at 20 °C, SAXS at 20 °C, DLS at 20 °C were performed to verify and model the homodimer. Irrespective of the methods and their working temperatures, the physical parameters like molecular weight (MW), the radius of gyration (Rg), the radius of hydration (Rh), estimated from the different methods have all converged (Table S6). Even the geometry of the dimer estimated from the Rg/Rh values from different techniques predicted a nonspherical geometry, as obtained from the SAXS (Table S7). Molecular details from in silico MD studies further revealed two distinct interfaces in the homodimer: electrostatic interactions between EC1 and EC1 and hydrophobic interactions between EC1 and EC2 of the opponent partners. The extended overlap of domains may impose stronger adhesion as observed in cellulo than the single domain overlap in classical cadherins. In classical cadherins, the thermodynamically stable trans-dimer is formed by the exchange of the N-terminal β-strands of the distal domain alone.

To quantify the adhesive strength of Cdh23 homodimers, we measured the lifetime (t₀ = 1/kₜₒₓ) of the complex from SMFS experiments using AFM. Two binding conformations were identified: one with interface one or EC1 alone, at a lower force range but with higher binding probability, and the other at a higher force range comprising both interfaces. The strength for interface one was weak (t₀ = 1.86 s) and comparable to that of classical cadherins (Table S8). The strength of the final structure (t₀ = 1224 s) is among the strong interactions between cadherins. The longer lifetime of the trans-dimer of Cdh23 was expected from the extended binding interface. The long lifetime of the Cdh23 homodimer thus clarifies the strong aggregation index measured for cells expressing Cdh23 than the cells adhered via classical E-cadherin. However, the Kₐ measured for Cdh23 is in the micromolar range, comparable to the classical cadherins (Table S8). The similar trend was also observed for nonclustered protocadherins (protocadherin-19, for example), which too forms homodimers with extended overlap between multiple EC domains with Kₐ in micromolar ranges [2]. The disparity in the Kₐ values with the area of the interacting interface may be attributed to the slow on-rate (~10⁴ M⁻¹s⁻¹) of the interactions which can be computed from the ratio of the off-rate and Kₐ.

The interface of Cdh23 mediated trans-homodimer is amphiphilic and involves residues that are conserved across a variety of species (Fig. S2). A missense mutation at the binding interface was identified in patients who have cutaneous cancer, indicating a physiological relevance of the interface. Further, the interface was validated by introducing a single point mutation (E78K), which impaired the dimer complex.

Deciphering the contribution of nonclassical cadherins in cell adhesion is gaining interests, especially the giant cadherins with long extracellular domains [42–44]. Primary focus on this area has been to identify nonclassical cadherin mediating cell adhesion junctions [45–47], understand their packing conformations at the junction [42,48], the function of the junction [44,49,50], and the molecular structure of the junction [8]. Deciphering the molecular structure of the trans-homodimer of Cdh23 is an essential and timely observation in this context, which may pave the way for understanding the molecular details of the strong cell adhesion junctions by atypical cadherins.

**Materials and methods**

**Cloning, expression, and purification**

Mouse Cdh23 (NP_075859.2) extracellular (EC) repeats EC1 (Q24 to D124) and EC1-2 (Q24 to D228) were PCR-amplified and cloned into pET21a vector (Novagen, Merck-Sigma, St. Louis, MO, USA). We considered the Q24 as Q2 in our experiments. Cdh23-EC1, EC1-2, and EC1-3 repeats were expressed in *E. coli* BL21CodonPlus (DE3)-RIPL cells (Stratagene, San Diego, CA, USA) and *E. coli* lemo(DE3) cultured in Luria/Bertani broth (Hi-Media, Mumbai, Maharashtra, India) and grown at 37 °C to an OD₆₅₀ of 0.6. Induction was carried out differently for both the strains. For expression in *E. coli* lemo(DE3), the cells were induced along with 1 mM rhamnose and 0.2 mM IPTG (Hi-Media) and then induced at 30 °C for 6 h. For expression in *E. coli* BL21CodonPlus (DE3)-RIPL, the protein was then induced with 0.2 mM IPTG and incubated at the same temperature. Unlike in *E. coli* BL21CodonPlus (DE3)-RIPL, Cdh23 EC1-2 protein from *E. coli* lemo(DE3) was obtained in the soluble fraction. For expressions in *E. coli* BL21CodonPlus (DE3)-RIPL, we revived the
proteins from the inclusion bodies and processed through serial dialysis, as reported [8]. All proteins were first purified using Ni-NTA beads followed by SEC on a Superdex 200 column (GE Healthcare, Chicago, IL, USA) in 25 mM HEPES (pH 7.5), 25 mM KCl, 100 mM NaCl, and 2 mM CaCl$_2$.

Analytical size exclusion chromatography

We injected 100 µM of purified Cdh23 EC constructs on Superdex 200 increase 10/300 column (GE Healthcare) at a flow rate of 0.3 mL·min$^{-1}$. Prior to the loading of the protein, the column was washed with degassed ultrapure water and equilibrated with degassed SEC buffer (25 mM HEPES, 25 mM KCl, 50 mM NaCl, 2 mM CaCl$_2$, pH 7.5). The column was calibrated using standard protein mix kit (Merck-Sigma) with proteins, cytochrome-C (12 000 Da), carbonic anhydrase (29 000 Da), albumin (66 000 Da), and alcohol dehydrogenase (76 000 Da). The standard calibration curve was generated by estimating the $K_v$ (partition coefficient) using the equation, $V'/V_0 = K_v$, where $V'$ is the elution volume ($V_0$) is void volume, 8 mL for a 24 mL ($V_0$) Sephadex 200 Increase size exclusion column. The eluted fractions for all the proteins were run on SDS/PAGE for molecular weight confirmation.

Single-molecule FRET measurements

For N-terminal labeling with dyes, we recombinantly modified valine at position 3 to cysteine (V3C) and attached to maleimide dyes (supplied by Lumiprobe) using the thiol-maleimide Michael addition reaction. The unreacted dye was removed using spin columns (10 kDa MWCO). The labeling ratio was measured from the absorbance at 280 nm for protein, 545 nm for Cy3, and 645 nm for Cy5. For C-terminal labeling, we followed sortase A (srtA)-mediated enzymatic reaction. For sortagging, we recombinantly modified the C-terminal of the proteins with LPETGGS. SrtA recognizes the sequence and inserts polyglycine by cleaving the T-G peptide bond. To introduce the dye, we used dye-modified polyglycine (GGGC-dye) [31].

smFRET measurements were performed using IX83 P2ZF inverted microscope (Olympus, Shinjuku, Tokyo, Japan) combined with IX3 TIR MITICO TIRF illuminator equipped with 532 nm diode laser system for cy3 excitation and 645 nm diode laser system for cy5 excitation. Fluorescence was collected using an oil-immersion objective (60X, NA 1.45, Olympus) into an EMCCD camera (Q-Imaging Roller Thunder, Surrey, BC, Canada). Image acquisition and processing were performed using CellSens Dimension (Olympus) software. iSMS software was used to localize the single-molecule dye pairs on the surface and then analyzed for their intensity profiles. The background subtraction and drift correction were also done using the same software. The FRET efficiency for each pair of donor-acceptor proteins was estimated using the same software. From the efficiency distributions, we estimated the distance between the FRET pairs using the following equation,

$$E = \frac{1}{1 + \left(\frac{R_0}{R}\right)^6}$$

For smFRET on surfaces, glass coverslips were freshly cleaned, silanized, pegylated, and finally modified with C-terminal of proteins specifically following a sortagging protocol as described elsewhere [31].

Live-cell binding and cell-cell aggregation assays

For the live-cell binding to surfaces, glass-bottom 96-well plates were cleaned, silanized, pegylated, and modified with proteins specifically. The C-terminal of the protein was attached covalently to the surfaces using sortagging as described before. Protein-coated surfaces were incubated with live A549 cells (10$^6$ cells) per well in 2 mM Ca$^{2+}$ buffer. After an incubation of 2 h, the surfaces were gently washed twice for 5 min with HEPES-Ca$^{2+}$ buffer and imaged for assessing cell density. The Ca$^{2+}$ from the surface was chelated by incubating with 1 mM EGTA, and again, the cell density was monitored. To number of live cells adhered to surfaces was quantified using MTT reagent. MTT reagent was added to each well and incubated for an hour. The cells were then lysed with DMSO, and the OD was monitored at 570 nm.

For cell-cell aggregation assay, the cells 36 h after the transfection with Cdh23 were counted and resuspended in HBSS buffer supplemented with Ca$^{2+}$ ions to a final cell count of ~10$^5$ cells. The aggregation was initiated by incubating the cells at 80 rpm and then imaged using bright field at 10× magnification using a Leica microscope.

SAXS data acquisition and analysis

The SAXS data were acquired for a q-range of 0.1 to 4 Å$^{-1}$ on a SAXSpace instrument (Anton Paar GmbH). The X-ray scattering setup had a slit-collimated X-ray source with a wavelength of 0.154 nm. The data were collected on a Mythen (Dectris, Baden-Daettwil, Switzerland) detector placed at a distance of 317.6 mm from the sample for 60 min (20 min × 3 frames). SAXStreat software was then used to calibrate the data for the beam position. The SAXSquant software was then used to subtract buffer contribution, set the usable q-range, and desmear the data using the beam profile. For each experiment, 100 µL of protein solution (Cdh23 EC1-2 and its mutant E78K) and their corresponding buffers were exposed to X-rays in a quartz capillary at a temperature of 10 °C. Data processing provided the scattering intensity (I) as a function of momentum transfer vector $q$ ($q = 4\pi\sin(\theta)/\lambda$, where $\theta$ and $\lambda$ are the...
scattering angle and the X-ray wavelength, respectively). The normalized Kratky plots ($I(q) \times q^2$ vs. $q \times R_g$) were made from the SAXS data using the program SCAFF (http://www.bioisis.net/) to interpret whether the protein remains folded during the SAXS data collection. The Guinier approximation was carried out using PRIMUS [24] of ATSAS 2.7 suite of programs [51] to estimate the radius of gyration ($R_g$) of the major scattering species.

Using GNOM program [26], we carried out the indirect Fourier transformation of SAXS data to obtain the probability distribution of the pairwise vectors ($P(r)$ curve) arising from scattering of the protein molecule in solution. The $P(r)$ curve analysis was done to acquire the maximum linear dimension ($D_{max}$) and the $R_g$ in real space.

**Shape reconstruction**

Ten independent models were generated using DAMMIF [27] program. The models were aligned, averaged, and filtered using DAMAVER [52] suite of programs. The averaged envelope was further refined using DAMMIN [28] program. This procedure provided an envelope that reflected the shape of a protein molecule in solution.

**Protein docking**

PatchDock is an online tool that follows rigid-body docking optimization with shape complementarity and two-point interactions between hot-spots. Hot-spots are decided based on residues that are conserved in protein–protein interaction surfaces and mediate salt-bridge type interactions, H-bonding, hydrophobic interactions, aromatic pi-stacking etc.

The available crystal structure of monomer Cdh23 EC1-2 was docked using PatchDock server [53] to generate the homodimer. Z-test was performed using SASSA-based parameters like linear dimension and $R_g$ and FRET-based end-to-end distances between the docked models and SAXS-based envelope (Table S7). The structures which agreed the most with the SAXS-based envelope of the protein remains folded during the SAXS data collection, and the emission monochromator for tryptophan was fixed at 342 nm, at a slit width of 8 nm. The instrument response function was measured using 2% LUDOX (Sigma-Aldrich).

**MD simulations**

MD simulations were performed on the in-house workstation. The crystal structure of the docked model which agreed the most with the SAXS-based envelop was used for simulations. Simulations were performed with GROMACS 5.0.1 using All-atom OPLS force field and TIP4P water model. All analyses were performed using VMD. For MD simulations, the Cdh23-dimer model was placed at the center of a $13.6 \times 5.7 \times 5.8$ nm triclinic box filled with four-point charge water molecules such that no atom of the protein was closer than 1 nm from the walls of the box. The MD equilibration system consisted at an average of 54 435 atoms. System charge neutrality of the system was maintained by adding Na$^+$ counterions to the box as needed with buffer ions (50 mM NaCl, 50 mM KCl, and 2 mM CaCl$_2$). To ensure that the solvated Cdh23-dimer model has no steric clashes or inappropriate geometries, we first energy minimized the system. Periodic boundary conditions were assumed in all simulations. A cutoff of 1 nm was used for van der Waals interactions. Electrostatic interactions were calculated with a particle mesh technique for Ewald summations with a 1 nm cutoff. We equilibrated the water molecules and ions around the Cdh23-dimer model in two phases. In the canonical (NVT) phase, the system was established at a constant reference temperature of 300 K. The pressure of the system was then stabilized under isothermal-isobaric (NPT) conditions. Following equilibration, 100 ns MD simulations were run at 2 fs integration steps and frames were recorded at 1-ps interval. We repeated the simulations 3 times.
times with the same docked structure and obtained overlapping results. We used visual molecular dynamics (VMD) for H-bond analysis.

**Analytical ultracentrifugation**

Analytical ultracentrifugation (AUC) experiments were carried out using a Beckman XLA/I ultracentrifuge and equipped with a Ti50An rotor using 12-mm six-channel cell centerpieces with sapphire windows and detection by UV at 280 nm. Both sedimentation velocity and equilibrium experiments were performed at 20 °C at pH 7.6 buffer containing 25 mM HEPES, 50 mM NaCl, 25 mM KCl, and 2 mM CaCl2. Prior to each experiment, the protein sample was dialyzed in the buffer for 18 h. Sedimentation velocity experiments were performed for all proteins at a rotor speed of 142 000 g. 300 scans were taken consecutively. Data were analyzed using SEDFIT software following continuous C(s) distribution, C(M) distribution, and C(s), (f/f0) models. For equilibrium, samples were subjected to fast spins at 99 000 g for 20 h to achieve rapid equilibrium. Then, we took one scan after each one-hour interval for 5 h and checked the rmsd fluctuations to test whether samples had reached equilibrium or not. Then, we decreased the rotor speed to 32 000 rpm and wait for 4 h before taking 3 scans consecutively. The same procedure has been followed for two other speeds 28 000 and 24 000 rpm, respectively. Buffer viscosity and density were measured using SEDNTERP (http://sednterp.unh.edu/). SEDPHAT was used to estimate the dissociation constant. We performed global fitting with mass conservation following monomer–dimer association model keeping baseline, meniscus, bottom, binding affinity as floating parameters.

**SMFS experiments**

For SMFS using AFM (Nano Wizard 3, JPK Instruments, Berlin, Germany), the protein molecules were immobilized on the glass coverslip and Si3N4 cantilever (Olympus, OMCL-TR400PSA-1), using a high-specificity immobilization protocol as described before [31].

For quantitative estimation of force from each experiment, the spring constant of the cantilever was measured using SEDNTERP (http://sednterp.unh.edu/). SEDPHAT was used to estimate the dissociation constant. We performed global fitting with mass conservation following monomer–dimer association model keeping baseline, meniscus, bottom, binding affinity as floating parameters.

The analyses for the plotted force curves were done in MATLAB with home-written programs. The single-molecule events were selected from the fit to freely joint chain (FJC) model using the following equation:

\[
\frac{1}{v_F} = \frac{1}{k_c v} + \frac{1}{v} \frac{d(l(F))}{dF} = \frac{1}{k_c v} \left[ 1 + \frac{k_c L_0}{k_BT} \left( \frac{k_BT}{k_BT} - \frac{2}{\cosh^2 \left( \frac{Fa}{k_BT} \right)} \right) \right]
\]

where \(x\) is Kuhn length, \(L_c\) is contour length (CL), and \(l(F)\) is stretching of PEG at every force. The unbinding forces were estimated from the peak maxima of the single-molecule unbinding events for each loading rate and plotted as distribution. Bin width was estimated from Scott’s method [57]. Each force distribution was hence fitted to Gaussian distribution, and the most probable force \((F)\) for each loading rate was obtained from the fit. Loading rate at each velocity was calculated using the following equation [39]

\[
F(v_F) = \left( \frac{k_BT}{x_F} \right) \ln \frac{v_F x_F}{\frac{1}{k_visitor}, \frac{1}{k_observer}} k_BT
\]

Non-specific binding rates were estimated in two different ways: by modifying either of the two surfaces (the cantilever or the coverslip) identically but without attaching proteins or by performing the same single-molecule protein–protein unbinding experiments in the absence of Ca\(^{2+}\) ions in the Chelex buffer and EGTA buffer. In both cases, the non-specific events accounted for < 0.5% of the total number of total selected PEG stretching events.

**Acknowledgements**

This work was supported by the Wellcome Trust/DBT India Alliance Fellowship [grant number: IA/1/15/1/501817] awarded to SR. We thank Professor James W. Nelson, Stanford University, and Professor William Weis from Stanford University for critically commenting on this manuscript which certainly enhanced the quality of the work. S.R. acknowledges the financial support by the Wellcome Trust/DBT Intermediate fellowship by India Alliance and Indian Institute of Science Education and Research Mohali (IISERM). G.S.S, A.S., and J.P.H sincerely thank IISERM for financial support. A.K. is grateful to DST INSPIRE...
for funding. J.S.S. is thankful to the Centre of Excellence (COE) in Frontier Areas of Science and Technology (FAST) program of the Ministry of Human Resource Development, Government of India for financial support. M.K.S. thanks the Wellcome Trust/DBT India Alliance.

Conflict of interests

The authors declare no conflict of interest.

Author contribution

SR has supervised the project. GSS and JSS did the cloning, expression, and purification. GSS, AK, AS, A, and SR recorded and analyzed SAXS. JPH recorded and analyzed AUC. GSS, AK, AS, RMY, and SR analyzed MD. GSS performed SMFS, smFRET, cell aggregation, CD, SDS/PAGE, WB. MKS, and GSS performed live-cell binding assay. GSS, AS, MKS, and JPH made the figures. SR and MKS RMY, and SR performed live-cell binding assay. AS, GSS, AK, JSS, JPH, AS, GSS, AS, MKS, and JPH made the figures. SR and MKS, and GSS performed live-cell binding assay. AS, GSS, AK, JSS, JPH, AS, RMY, A, and SR recorded and analyzed SAXS. JPH cloning, expression, and purification. GSS, AK, AS, SR has supervised the project. GSS and JSS did the cloning, expression, and purification. GSS, AK, AS, SR performed live-cell binding assay. AS, GSS, AK, AS, RMY, A, and SR recorded and analyzed SAXS. JPH cloning, expression, and purification. GSS, AK, AS, SR recorded and analyzed SAXS.

References

1 Gul IS, Hulpiua P, Saeys Y & van Roy F (2017) Evolution and diversity of cadherins and catenins. Exp Cell Res 358, 3–9.
2 Cooper SR, Jontes JD & Sotomayor M (2016) Structural determinants of adhesion by proteocadherin-19 and implications for its role in epilepsy. eLife 5, 1–22.
3 Tanoue T & Takeichi M (2005) New insights into Fat cadherin adhesion. J Cell Sci 118, 2347–2353.
4 Uhlen M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjöstedt E & Asplund A et al. (2015) Proteomics. Tissue-based map of the human proteome. Science 347, 1260419.
5 Kazmierczak P, Sakaguchi H, Tokita J, Wilson-Kubalek EM, Milligan RA, Müller U & Kachar B (2007) Cadherin 23 and protocadherin 15 interact to form tip-link filaments in sensory hair cells. Nature 449, 87–91.
6 Sannigrahi MK, Srinivas CS, Deokate N & Rakshit S (2019) The strong propensity of Cadherin-23 for aggregation inhibits cell migration. Mol Oncol 13, 1092–1109.
7 Cancer Genome Atlas Research Network (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 455, 1061–1068.
8 Sotomayor M, Weihofer WA, Gaudet R & Corey DP (2012) Structure of a force-conveying cadherin bond essential for inner-ear mechanotransduction. Nature 492, 128–132.
9 Bork JM, Peters LM, Riazuddin S, Bernstein SL, Ahmed ZM, Ness SL, Polomero R, Ramesh A, Schloss M, Srissailpathy CR et al. (2001) Usher syndrome 1D and nonsyndromic autosomal recessive deafness DFN12 are caused by allelic mutations of the novel cadherin-like gene CDH23. Am J Hum Genet 68, 26–37.
10 Siemens J, Lillo C, Dumont R, Reynolds A, Williams DS, Gillespie PG & Müller U (2004) Cadherin 23 is a component of the tip link in hair-cell stereocilia. Nature 428, 950–955.
11 Sollner C, Rauch G-J, Siemens J, Geisler R, Schuster SC, Müller U & Nicolson T (2004) Mutations in cadherin 23 affect tip links in zebrafish sensory hair cells. Nature 428, 955–959.
12 Lagziel A, Ahmed ZM, Schultz JM, Morell RJ, Belyantseva IA & Friedman TB (2005) Spatiotemporal pattern and isoforms of cadherin 23 in wild type and walzter mice during inner ear hair cell development. Dev Biol 280, 295–306.
13 Brasch J, Harrison OJ, Honig B & Shapiro L (2012) Thinking outside the cell: How cadherins drive adhesion. Trends Cell Biol 22, 299–310.
14 Sivasankar S, Zhang Y, Nelson WJ & Chu S (2009) Characterizing the initial encounter complex in cadherin adhesion. Structure 17, 1075–1081.
15 Rakshit S, Zhang Y, Manibog K, Shafraz O & Sivasankar S (2012) Ideal, catch, and slip bonds in cadherin adhesion. Proc Natl Acad Sci USA 109, 18815–18820.
16 Manibog K, Sankar K, Kim S-A, Zhang Y, Jernigan RL & Sivasankar S (2016) Molecular determinants of cadherin ideal bond formation: conformation-dependent unbinding on a multidimensional landscape. Proc Natl Acad Sci USA 113, E5711–E5720.
17 Garrod DR, Berika MY, Bardsey WF, Holmes D & Taberner L (2005) Hyper-adhesion in desmosomes: its regulation in wound healing and possible relationship to cadherin crystal structure. J Cell Sci 118, 5743–5754.
18 Ciatto C, Bahna F, Zampieri N, VanSteenhouse HC, Katsamba PS, Ahslen G, Harrison OJ, Brasch J, Jin X, Posy S et al. (2010) T-cadherin structures reveal a novel adhesive binding mechanism. Nat Struct Mol Biol 17, 339–347.
19 Sotomayor M, Weihofer WA, Gaudet R & Corey DP (2010) Structural determinants of cadherin-23 function in hearing and deafness. Neuron 66, 85–100.
20 Srinivasan S, Hazra JP, Singaraju GS, Deb D & Rakshit S (2017) ESCORTing proteins directly from whole cell-lysate for single-molecule studies. Anal Biochem 535, 35–42.
21 Garg S, Singaraju GS, Yengkhom S & Rakshit S (2018) Tailored polyproteins using sequential staple and cut. Bioconjug Chem 29, 1714–1719.
22 Preus S, Noer SL, Hildebrandt LL, Gudnason D & Birkedal V (2015) iSMS: single-molecule FRET microscopy software. Nat Methods 12, 593–597.
23 Shan W, Yagita Y, Wang Z, Koch A, Svenningsen AF, Gruzglin E, Pedraza L & Colman DR (2004) The minimal essential unit for cadherin-mediated intercellular adhesion comprises extracellular domains 1 and 2. J Biol Chem 279, 55914–55923.
24 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 Crystallogr 36, 1277–1282.
25 Rambo RP & Tainer JA (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. Biopolymers 95, 559–571.
26 Svergun DI (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J Appl Crystallogr 25, 495–503.
27 Franke D & Svergun DI (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. J Appl Crystallogr 42, 342–346.
28 Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys J 76, 2879–2886.
29 Schuck P (2003) On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. Anal Biochem 320, 104–124.
30 Katsamba P, Carroll K, Ahlsen G, Bahna F, Vendome J, Posy S, Rajebhosale M, Price S, Jessell TM, Ben-Shaul A et al. (2009) Linking molecular affinity and cellular specificity in cadherin-mediated adhesion. Proc Natl Acad Sci USA 106, 11594–11599.
31 Vendome J, Felsovalyi K, Song H, Yang Z, Jin X, Brash J, Harrison OJ, Ahlsen G, Bahna F, Kaczynska A et al. (2014) Structural and energetic determinants of adhesive binding specificity in type I cadherins. Proc Natl Acad Sci USA 111, E4175–E4184.
32 Brown PH, Balbo A & Schuck P (2008) Characterizing protein-protein interactions by sedimentation velocity analytical ultracentrifugation. Curr Protoc Immunol 81, 1–39.
33 Zhao H, Berger AJ, Brown PH, Kumar J, Balbo A, May CA, Casillas E, Laue TM, Patterson GH, Mayer ML & et al. (2012) Analysis of high-affinity assembly for AMPA receptor amino-terminal domains. J Gen Physiol 139, 371–388.
34 Göhler A, Büchner C, André S, Doose S, Kaltner H & Gabius HJ (2011) Sensing ligand binding to a clinically relevant lectin by tryptophan fluorescence anisotropy. Analyst 136, 5270–5276.
35 Montecinos-Franjola F, Schuck P & Sackett DL (2016) Tubulin dimer reversible dissociation. J Biol Chem 291, 9281–9294.
36 Merrin J, Kumar P & Libchaber A (2011) Effects of pressure and temperature on the binding of RecA protein to single-stranded DNA. Proc Natl Acad Sci USA 108, 19913–19918.
37 Srivivasan S, Hazra JP, Singaraju GS, Deb D & Rakshit S (2017) ESCORTing proteins directly from whole cell-lysate for single-molecule studies. Anal Biochem 535, 35–42.
38 Marshall L, Helgadóttir H, Mölle M & Born J (2006) Boosting slow oscillations during sleep potentiates memory. Nature 444, 610–613.
39 Ray C, Brown JR & Akhremtchev BB (2007) Rupture force analysis and the associated systematic errors in force spectroscopy by AFM. Langmuir 23, 6076–6083.
40 Bell GI (1978) Models for the specific adhesion of cells to cells. Science 200, 618–627.
41 Evans E & Ritchie K (1997) Dynamic strength of molecular adhesion bonds. Biophys J 72, 1541–1555.
42 Tsukasaki Y, Miyazaki N, Matsumoto A, Nagae S, Yonemura S, Tanoue T, Iwasaki K & Takeichi M (2014) Giant cadherins Fat and Dachsous self-bend to organize properly spaced intercellular junctions. Proc Natl Acad Sci USA 111, 16011–16016.
43 Goodman KM, Rubinstein R, Thu CA, Mannepalli S, Bahna F, Ahlsen G, Rittenhouse C, Maniatis T, Honig B & Shapiro L (2016) γ-Protocadherin structural diversity and functional implications. eLife 5, 1–25.
44 Caldwell J, Collins M, Kemke K, Sherrill J, Wen T, Rochman M, Stucke E, Amin L, Tai H, Putnam P et al. (2016) Cadherin 26 is an alpha integrin-binding epithelial receptor regulated during allergic inflammation. Mucosal Immunol 10, 1190–1201.
45 Sannigrahi MK, Srivinas S & Rakshit S (2018) The prospects of cadherin-23 as a mediator of homophilic cell-cell adhesion. In Biochemical and Biophysical Roles of Cell Surface Molecules (Chattopadhyay K & Basu SC, eds), pp. 99–105. Springer Singapore, Singapore.
46 Van Roy F (2014) Beyond E-cadherin: roles of other cadherin superfamily members in cancer. Nat Rev Cancer 14, 121–134.
47 Nakajima D, Nakayama M, Kikuno R, Hiroswa M, Nagase T & Ohara O (2001) Identification of three novel non-classical cadherin genes through comprehensive analysis of large cDNAs. Mol Brain Res 94, 85–95.
48 Tariq H, Bella J, Jowitt TA, Holmes DF, Rouhi M, Nie Z, Ballock C, Garrod D & Tabernero L (2015) Cadherin flexibility provides a key difference between desmosomes and adherens junctions. Proc Natl Acad Sci USA 112, 5395–5400.
49 Langhe RP, Gudzenko T, Baehmann M, Becker SF, Gonnemann C, Winter C, Abbruzzese G, Alfandari D,
Kratzer MC, Franz CM & et al. (2016) Cadherin-11 localizes to focal adhesions and promotes cell-substrate adhesion. *Nat Commun* 7, 1–10.

Goodman KM, Rubinstein R, Dan H, Bahna F, Mannepalli S, Ahsén G, Aye Thu C, Sampogna RV, Maniatis T, Honig B & et al. (2017) Protocadherin cis-dimer architecture and recognition unit diversity. *Proc Natl Acad Sci USA* 114, E9829–E9837.

Petoukhov MV, Franke D, Shkumatov AV, Tria G, Kikhney AG, Gajda M, Gorba C, Mertens HDT, Konarev PV & Svergun DI (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J Appl Crystallogr* 45, 342–350.

Volkov VV & Svergun DI (2003) Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Crystallogr* 36, 860–864.

Schneidman-Duhovny D, Inbar Y, Nussinov R & Wolfson HJ (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res* 33, 363–367.

Kozin MB & Svergun DI (2001) Automated matching of high- and low-resolution structural models. *J Appl Crystallogr* 34, 33–41.

Schrödinger LLC (2015) The [PyMOL] Molecular Graphics System, Version-1.8. Schrödinger, LLC, New York, NY.

Hutter JL & Bechhoefer J (1993) Calibration of atomic-force microscope tips. *Rev Sci Instrum* 64, 1868–1873.

Scott DW (1979) On optimal and data-based histograms. *Biometrika* 66, 605–610.

**Supporting information**

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Fig. S1.** Multiple sequence alignment showing key residues driving homodimerization in type I and type II cadherins versus Cdh23

**Table S1.** End to end distances obtained from smFRET data analysis along with photo-physical properties of fluorophores.

**Table S2.** Data collection parameters are tabulated along with the software used for analysing the scattering data.

**Table S3.** Determination of the rotational correlation decay time of Trp66 (W66) at three different concentrations of Cdh23 EC1-2 WT.

**Table S4.** Quantitative comparison of different PatchDock structures.

**Table S5.** Parameters used during MD simulations.

**Table S6.** Physical parameters of the trans-homodimer of Cdh23 EC1-2 (WT), estimated from various techniques.

**Table S7.** Estimation of $R_g/R_H$ of the trans-homodimer of Cdh23 EC1-2 (WT) obtained from various methods.

**Table S8.** A comparison of dissociation constants and off-rate values for first two domains of various cadherins.