

SCIENTIFIC REPORTS



OPEN

EpCAM homo-oligomerization is not the basis for its role in cell-cell adhesion

Aljaž Gaber¹, Seung Joong Kim², Robyn M. Kaake³, Mojca Benčina⁴, Nevan Krogan^{3,5}, Andrej Šali², Miha Pavšič¹ & Brigita Lenarčič^{1,6}

Cell-surface tumor marker EpCAM plays a key role in proliferation, differentiation and adhesion processes in stem and epithelial cells. It is established as a cell-cell adhesion molecule, forming intercellular interactions through homophilic association. However, the mechanism by which such interactions arise has not yet been fully elucidated. Here, we first show that EpCAM monomers do not associate into oligomers that would resemble an inter-cellular homo-oligomer, capable of mediating cell-cell adhesion, by using SAXS, XL-MS and bead aggregation assays. Second, we also show that EpCAM forms stable dimers on the surface of a cell with pre-formed cell-cell contacts using FLIM-FRET; however, no inter-cellular homo-oligomers were detectable. Thus, our study provides clear evidence that EpCAM indeed does not function as a homophilic cell adhesion molecule and therefore calls for a significant revision of its role in both normal and cancerous tissues. In the light of this, we strongly support the previously suggested name Epithelial Cell Activating Molecule instead of the Epithelial Cell Adhesion Molecule.

Epithelial Cell Adhesion Molecule (EpCAM) is a cell-surface type I transmembrane glycoprotein, which was initially identified as a colorectal carcinoma antigen^{1,2}. Due to its elevated expression levels on carcinoma cells³ it is widely accepted as a target for drug delivery as well as a molecular marker in tumor diagnostics and prognostics (reviewed in ref.⁴). Here, high EpCAM expression is often associated with poor prognosis^{5–11} and linked to cancer proliferation, migration and metastasis¹². Besides this, EpCAM expression is also high on undifferentiated human embryonic stem cells^{13–15}. The molecular mechanisms on which EpCAM's function in normal and cancerous tissue is based are still not explained completely. However, two major roles have been described – cell proliferation-enhancing signaling, which has been extensively studied in recent years^{16–18}, and cell-cell adhesion, that hasn't received abundant attention since its first description more than two decades ago¹⁹ and which we are addressing here in detail.

Signaling via EpCAM involves regulated intermembrane proteolysis (RIP) (Fig. 1a), resulting in shedding of EpCAM's extracellular domain (EpEX). Additional cleavages within the transmembrane region lead to release of the intracellular domain (EpIC) that is recruited into formation of EpIC-FHL2-Lef1-β-catenin complex that eventually binds to promoter region in the nucleus^{16,20}. Through the interaction with Lef1, the EpIC-containing nuclear complex regulates the expression of proliferating factors such as cyclin D1 and *c-myc*^{16,21}. RIP, the initial event in signaling, is believed to be triggered by *trans*-oligomerization of EpCAM's extracellular domain, *i.e.* by binding of free EpEX to cell-bound EpCAM or via inter-cellular interaction between EpCAM molecules on adjacent cells²², as opposed to *cis*-oligomerization involving EpCAM molecules on the same cell. Alternatively, yet unknown ligand(s) of EpCAM may be instrumental in initiating RIP.

¹Department of Chemistry and Biochemistry, Faculty of Chemistry and Chemical Technology, University of Ljubljana, Večna pot 113, Ljubljana, SI 1000, Slovenia. ²Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, California Institute for Quantitative Biosciences, University of California, San Francisco, 1700 4th Street, Suite 503B, San Francisco, CA, 94158, USA. ³J. David Gladstone Institutes, San Francisco, CA, 94158, USA. ⁴Department of Synthetic Biology and Immunology, National Institute of Chemistry, Hajdrihova 19, Ljubljana, SI 1000, Slovenia. ⁵Quantitative Biosciences Institute, QBI, Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, CA, 94158, USA. ⁶Department of Biochemistry, Molecular and Structural Biology, Institute Jožef Stefan, Jamova 39, Ljubljana, SI 1000, Slovenia. Correspondence and requests for materials should be addressed to M.P. (email: miha.pavsic@fkkt.uni-lj.si) or B.L. (email: brigita.lenaric@fkkt.uni-lj.si)

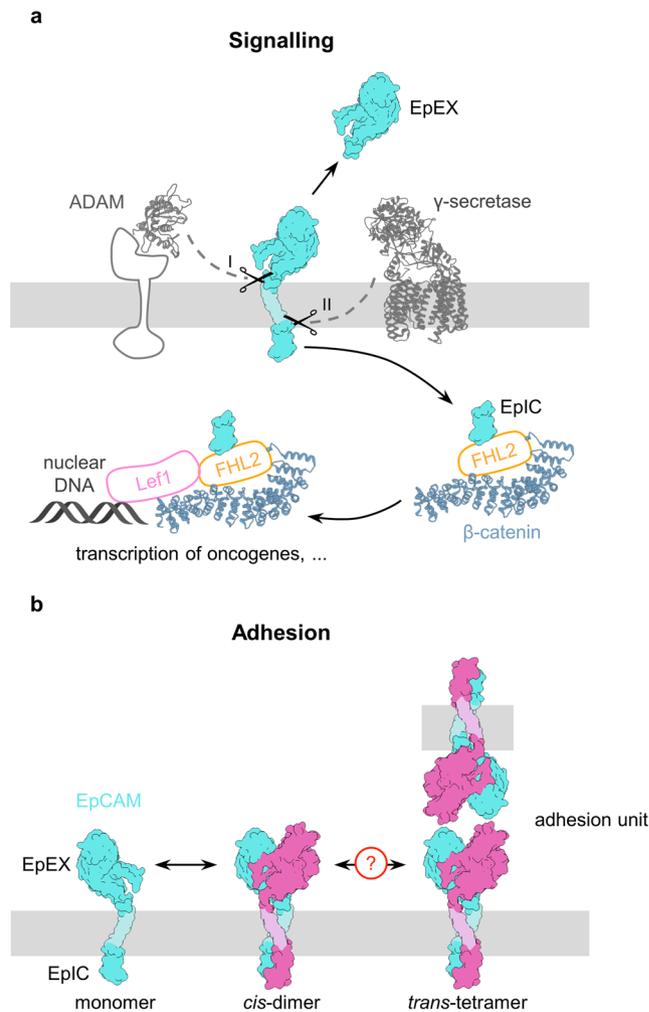


Figure 1. EpCAM's function in cell-cell adhesion and signalling. **(a)** EpCAM mediated adhesion. EpCAM is depicted as a shape outline. Subunits in *cis*-dimers are colored cyan and magenta, transmembrane regions are depicted in paler colors. Cell membrane is shown in grey. **(b)** EpCAM signalling via RIP. EpCAM is depicted as a shape outline. Structures of ADAM's catalytic domain, γ -secretase, β -catenin are presented as ribbons (PDB: 1bkc, 5a63 and 2z6h, respectively). The portion ADAM beside the catalytic domain, FHL2 and Lef1, whose 3D structures are not known are depicted as shapes with sizes corresponding to their mass.

Oligomerization is also the principal feature of adhesion complexes formed during the establishment of proposed cell-cell contacts *via* EpCAM (Fig. 1b). Two models of inter-cellular oligomerization were proposed: formation of *trans*-octamers from *cis*-tetramers²³ and, much later, formation of *trans*-tetramers from *cis*-dimers on adjacent cells²⁴. The latter model is supported by the solved crystal structure of EpEX dimer²⁵ in which the orientation of the two subunits is consistent with *cis*-orientation, and proposes a D2 symmetry of the *trans*-tetramer.

Initial discoveries describing EpCAM as a novel calcium independent homophilic cell adhesion molecule, including translocation to contact areas upon formation of cell-cell contacts²⁶, disruption of the latter by EpCAM targeting antibodies²⁶, and the ability to induce cell aggregation in transfected mouse fibroblast L929 cells¹⁹, were soon followed by indications that EpCAM's role in cell-cell adhesion is much more complex. For example, EpCAM has a negative effect on the strength of classical E-Cadherin mediated cell-cell adhesion^{27,28} and directly interacts with claudin-7, thus regulating the formation of functional tight junctions^{29,30}. Even more, despite notable evidence of EpCAM's ability to directly mediate cell-cell adhesion, there have been some contradicting reports. First, a year after initial description of EpCAM as a novel CAM Fornaro *et al.* failed to reproduce EpCAM-induced cell aggregation³¹. Most recently, Tsaktanis *et al.* showed neither EpCAM cleavage nor EpCAM knockdown have any effect on cell-cell adhesion in a carcinoma cell line²⁰. Furthermore, in our previous work with soluble EpEX²⁵ we also failed to observe any evidence of higher-order oligomerization (*i.e.*, more than dimerization). Combined, these findings encouraged us to reinvestigate the mechanism of EpCAM's role as a cell-cell adhesion molecule (CAM) on a molecular level with an aim to provide a detailed explanation of the observed inconsistencies regarding the protein's function.

Here we present structural analysis of EpCAM oligomerization and its ability to form higher-order oligomers, which would be consistent with the formation of homophilic cell-cell adhesion units. To gain insight

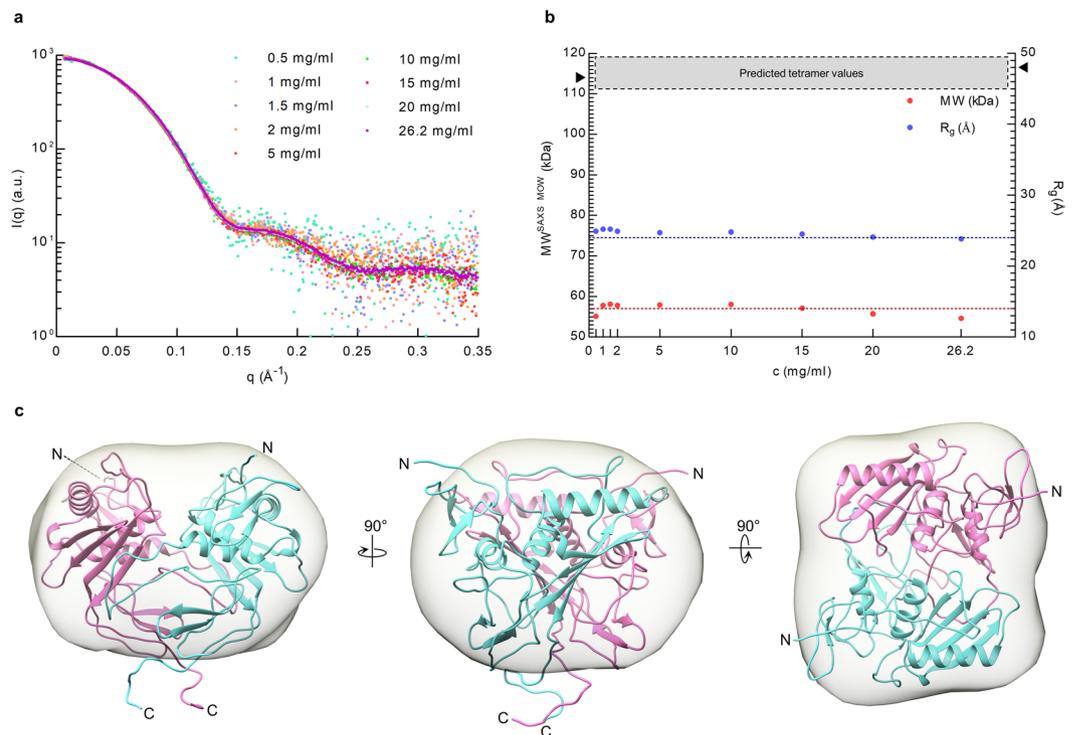


Figure 2. SAXS analysis shows no evidence of oligomerization. **(a)** Scaled SAXS profiles of ngEpEX at different concentrations in the range from 0.5 mg/ml (17.5 μ M) to 26.2 mg/ml (919.4 μ M) presented and overlaid in the same plot. **(b)** MW and R_g values calculated from SAXS profiles. Dotted lines represent values calculated for dimer structure 57 kDa and 24 Å. Predicted area of tetramer values are depicted with a grey rectangle and the hypothetical tetramer values at 114 kDa and 54 Å are marked with black triangles. **(c)** Three orientations of EpEX dimer structure docked in the *ab initio* shape (grey envelope) reconstructed from the merged SAXS profile. Subunits in the dimer are depicted as cyan and magenta ribbons.

into proteins behaviour both *in vitro* and *in vivo* we employed small angle X-ray scattering (SAXS), chemical cross-linking coupled with mass spectrometry (XL-MS), bead aggregation assay (BAA), and fluorescence-lifetime imaging based Förster resonance energy transfer (FLIM-FRET). Our data clearly demonstrate that while both EpCAM and EpEX form *cis*-dimers *in vitro* and *in vivo*, no notable higher-order oligomerization takes place. Even more, EpCAM molecules from adjacent cells do not form inter-cellular higher-order homo-oligomers, making EpCAM's function as a homophilic CAM highly implausible.

Results

EpEX *cis*-dimers do not form higher-order oligomers in solution. First we utilized SAXS on highly concentrated EpEX samples to see if high protein concentration leads to the formation of (significant) amount of higher-order oligomers which could be below the detection limit at low concentration, since the dissociation constant (K_d) of the proposed *trans*-tetramer is much higher than the K_d of the *cis*-dimer (estimated values of 10 nM and 10 μ M, respectively²⁴). Here, formation of tetramers would manifest itself most prominently as increase in radius of gyration (R_g), a measure of the average particle size. To reduce ambiguities in data interpretation and modeling, we utilized mutant non-glycosylated EpEX (ngEpEX) to avoid potential differences in glycosylation pattern in *wt* EpEX. Similarly, due to homogeneity issues we didn't use full length EpCAM (EpFL) embedded in detergent micelles. Still, our results should be applicable to *wt* EpCAM since glycosylation reportedly does not interfere with EpCAM *trans*-oligomerization²³ which is mediated exclusively via EpEX^{23,25}.

SAXS data were collected at several protein concentrations ranging from 0.5 to 26.2 mg/ml (corresponding to 17.5–919.4 μ M). At the highest sample concentration, approx. 93% of ngEpEX would be in tetrameric form considering a K_d of 10 μ M and a simple dimer \leftrightarrow tetramer equilibrium relation, *i.e.* an amount that could most certainly be detected as changes in R_g and D_{max} (maximum particle size) values by SAXS analysis³². Surprisingly, we observed no significant concentration dependent changes, as the scaled SAXS profiles at different sample concentration nearly coincided (Fig. 2a). Furthermore, SAXS profile-derived MW and R_g values, both indirectly describing average particle size, correspond well to the values calculated from the slightly modified dimer crystal structure²⁵ (PDB: 4MZV; C-terminal stretch of residues 259–265 was modelled as flexible). This indicates that the dimer is the highly predominant, if not the only oligomeric state of EpEX in the analyzed concentration range (Fig. 2b; Supplementary Table 1). This observation is further supported by the good size and shape agreement of the dimer X-ray structure with the *ab initio* shape reconstructed from the experimental SAXS profile (Fig. 2c).

To ensure radiation-preventing additives had no effect on the observed results, we collected SAXS profiles using two additional buffers (one buffer without glycerol, and another buffer containing 1% (w/v) sucrose and

5 mM Na/K Nitrate). While we did observe minor radiation damage in samples without any additives, SAXS profiles were highly consistent and comparable regardless of the buffer composition (Supplementary Table 1b,c).

Consistently, Multi-Fast open-source X-ray scattering (Multi-FoXS) analysis^{33,34} suggests that the population of the proposed *trans*-tetramer form is simply negligible, in all SAXS concentrations between 0.5 to 26.2 mg/ml. We first searched multi-state models from a combined ensemble of 10,000 dimer models (generated using Modeller³⁵ from the crystal structure of PDB 4MZV by accounting for flexible C-terminus) and 44,098 tetramer models (generated using Patchdock^{36,37} to obtain all possible configurations of two dimer units forming a tetramer). Next, we compared the calculated SAXS profiles of the multi-state models with the experimental SAXS data by using the χ^2 score value. The best fit of the resulting multi-state models did not significantly improve over the single-state dimer structures (Supplementary Fig. 1). The single-state dimer structures are highly consistent with the experimental SAXS data up to 15 mg/ml (χ^2 values of 1.39–1.97, Supplementary Fig. 1). Although minor inter-molecular repulsion was observed at higher protein concentrations (>15 mg/ml), it does not change any conclusion from the Multi-FoXS analysis. All resulting multi-state models that did improve the fit were solely comprised of dimer structure models with minor conformational rearrangements. In contrast, the tetramer structure models were completely excluded from the 100 best scoring multi-state models at all concentrations.

Collecting additional data at even higher protein concentrations than 26.2 mg/ml was not feasible since we observed the inter-molecular repulsion already at ~15 mg/ml. The repulsion is strongly indicated by smaller protein MW and R_g values (Supplementary Table 1a), supporting that at the high concentration ngEpEX dimers tend to repel each other rather than attract to form oligomers. All these combined results thus clearly suggest that EpEX dimers do not interact to form higher-order oligomers in solution.

Assignment of chemical cross-links as input data for structural modeling. Appearance of bands corresponding to the molecular mass of a dimer and tetramer in chemically cross-linked EpCAM samples is considered one of the fundamental proofs of EpCAM's adhesion-mediating oligomerization^{23,24,38}. To see if such data can be used as distance restraints in structural modeling of the proposed tetramer we used chemical cross-linking coupled with mass spectrometry (XL-MS) to identify the cross-links within electrophoretically-separated bands corresponding to the different oligomeric forms of EpCAM. This approach would at the same time enable us to capture transient oligomers that could be too short lived to be detected by SAXS. For these experiments we used both ngEpEX as well as ngEpFL thus addressing the observation that transmembrane and intracellular part are necessary for formation of higher-order oligomers²⁴. *Wt* glycosylated EpCAM was not used to avoid issues with heterogenic glycosylation during cross-link assignment by MS.

In agreement with previously published findings^{23,24}, four distinct bands were observed after SDS-PAGE analysis of cross-linked samples. Their masses correspond to mono-, di-, tri- and tetrameric ngEpEX or ngEpFL (Fig. 3a). Cross-linking was more efficient in the case of ngEpFL, resulting in higher percentage of higher-oligomeric forms compared to ngEpEX where the most intense band corresponds to a dimer (regardless of cross-linker concentration). Separate bands were excised, in gel digested, and resulting peptides separated and measured by LC-MS/MS. Using the cross-linked peptide analysis feature of Protein Prospector³⁹ we identified 21 unique cross-links (Fig. 3b). For structural modeling, we considered only the 18 cross-links which had both end-point residues in the ectodomain (cross-links 77–299, 129–296 and 129–299 were excluded since they involved the intracellular region of ngEpFL). Two cross-links were obtained exclusively from bands corresponding to a monomer, four to a dimer, and four to a higher-order oligomer, while the rest ($n = 8$) were identified from bands corresponding to more than one oligomeric state. Surprisingly, we did not observe any cross-links exclusively within the IC-tail, despite the abundance of lysines (5 out of 23 residues).

Unambiguous assignment of cross-links as intra- or inter-subunit within the various oligomeric states is not always possible, regardless of the band of origin on SDS-PAGE. To avoid false assignment, we assumed that all obtained cross-links can belong to any oligomeric form. The only exception were cross-links obtained exclusively from monomeric bands; these were postulated to be intra-subunit since they would otherwise cause significant shift in electrophoretic mobility (cross-links 61–218 and 65–77). All other cross-links were assigned to the oligomeric state in which they had the highest “Matched and Non-accessible cross-link score” (MNXL score) calculated from their Solvent Accessible Surface Distances (SASDs)⁴⁰. MNXL score distinguishes between experimentally observed cross-links, which have SASD lower than empirically determined threshold for DSS cross-linker (33 Å) – “matched cross-links” and all those who have larger SASDs or aren't present on the solvent accessible surface of the protein/protein complex – “non-accessible cross-links”. Scoring based on SASDs was used to avoid false identification of cross-links as intra-subunit as is often the case with the more commonly used Euclidean distances.

First, we calculated SASD for each cross-link and assigned them according to the published parameters using an adapted version of Jwalk⁴⁰ (see Materials and Methods for details). Distances were calculated for 17 cross-links in the subunit structure (total dataset excluding 129–129, which can't arise from an intra-subunit cross-link), and 16 cross-links in the dimer structure (total dataset excluding 61–218 and 65–77, which appeared only in bands corresponding to a monomer). For monomer, 9 cross-links could be matched while 8 had SASD larger than 33 Å (Fig. 3c; Supplementary Fig. 2). To account for monomer flexibility on a small scale (sidechains plus limited main chain conformational changes), equivalent SASDs derived from 1000 frames in 10 ns MD simulation (one frame per 10 ps) were calculated. Of the 8 non-accessible, two more (65–77 and 155–168) could be matched and two had distances just above the threshold of 33 Å (61–218 and 77–151) (Supplementary Fig. 2). In dimer structure, considering both intra- and inter-subunit distances, 12 of 16 cross-links were matched while 4 SASDs were longer than the threshold (Fig. 3d; Supplementary Fig. 2). They correspond to cross-links 77–168, 129–151, 155–168 and 155–218.

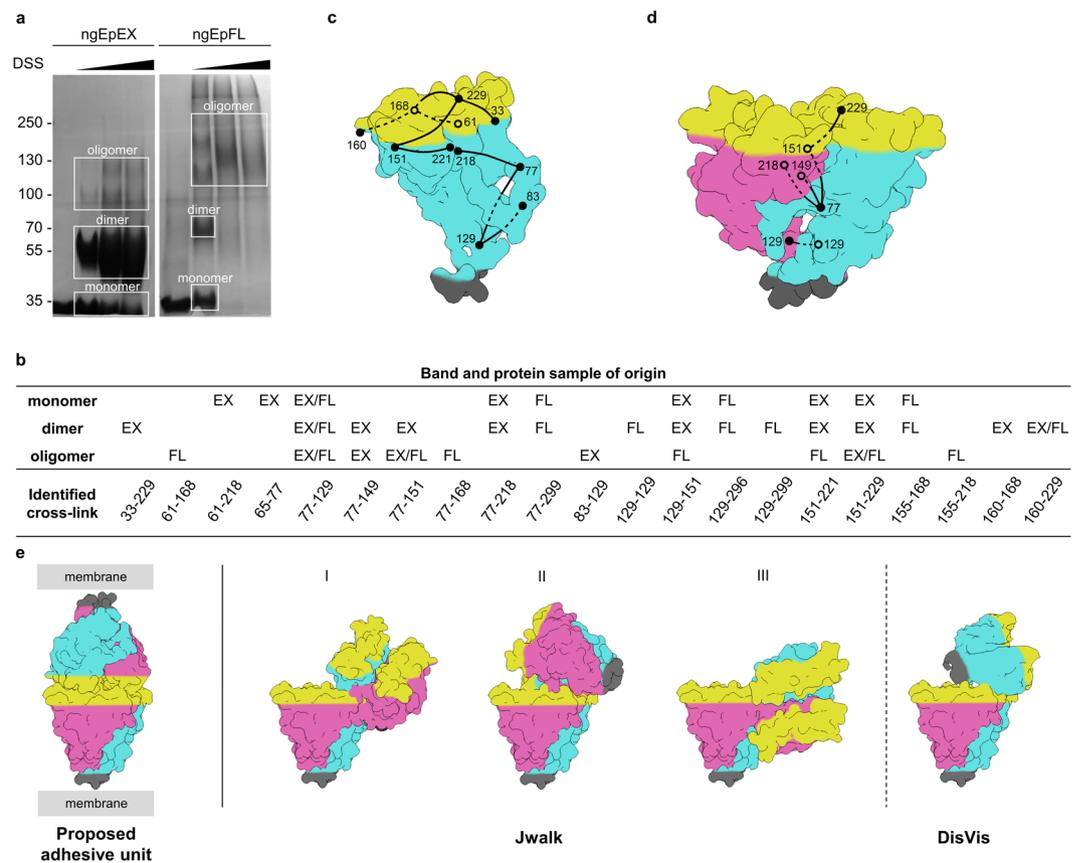


Figure 3. XL-MS analysis of EpCAM oligomerization. **(a)** SDS-PAGE analysis of cross-linking experiment. Black triangle at the top indicates increasing molar ratio of crosslinker (DSS) vs protein. White rectangles denote excised areas, which were analysed with MS. Lanes were cropped and greyscaled for clarity. Full-length gel is presented in Supplementary Fig. 3. **(b)** Cross-links identified in MS analysis. Columns represent corresponding areas of origin, EX and FL stand for identification in ngEpEX or ngEpFL XL experiment, respectively. **(c,d)** Shortest SASD for each matched cross-link in monomer model and dimer structure. In dimer structure, only one of each ambiguous inter-dimer cross-links is represented. Subunits are depicted as shape outlines. Subunits in *cis*-dimers are colored cyan and magenta, with membrane proximal and membrane distal parts depicted in grey and yellow, respectively. **(e)** Modelling of tetramers based on cross-links, non-accessible in dimer. Proposed *trans*-tetrameric adhesive unit model was taken from literature (Pavšič, 2014). Jwalk results are presented as three clusters, generated from 25 best scoring random tetramer models, clustered at 10 Å. DisVis results represent average ligand occupancy at default cut-off, as outputted by the web server.

Cross-links do not support a functionally relevant *trans*-tetramer model. After cross-link assignment four aforementioned cross-links remained which could not be matched neither to monomer nor to the *cis*-dimer structure. Therefore, we assumed that they stem from a tetrameric form of ngEpEX/ngEpFL. To see if they could indeed support a functionally relevant *trans*-tetramer model we performed an exhaustive explorative search of virtually all possible tetramer models with an aim to match those four remaining cross-links. At the same time, we attempted to see whether alternative structural models improve MNXL scores of the already matched cross-links.

The first step was the generation of a large set of virtually all possible random tetramer models (using a dimer crystal structure). This approach ensures that all possible dimer-dimer orientations are tested for their consistency with the experimental data in contrast to using spatial restraints from cross-links to guide the sampling from the beginning. Next, inter-dimer SASDs were calculated for the 44,098 random tetramer models (which were also used for multi-FoXS analysis above) and 12 D2 symmetric *trans*-tetramer models and compared them to corresponding intra- and inter-subunit SASDs in the dimer. None of tetramer models significantly improved the results for any cross-link already matched in the dimer, however some tetramer models can be used to match all four remaining non-accessible cross-links (Supplementary Fig. 2b). Importantly, random tetramer models outperform models with D2 symmetry in terms of satisfying the cross-link-derived distance restraints. Best tetramer model with D2 symmetry, i.e. model with the highest MNXL score, matched only two of the four previously non-accessible cross-links (Supplementary Fig. 2b).

To assess their functional significance the top-scoring models were compared to a proposed adhesion unit with respect to C-termini orientation corresponding to the membrane anchor point. For clarity, they were first clustered considering their symmetry with a 10 Å RMSD-cutoff. In none of the three resulting clusters the two

dimers are oriented in a way that could even approximate a theoretically plausible *trans*-interaction between dimers from adjacent cells (Fig. 3e).

Similar results were obtained using DisVis^{41,42} which uses Euclidean rather than solvent accessible surface distances. DisVis also accounts for the possibility that some of the obtained cross-links are false positives. We performed the analysis with the four cross-links that were non-accessible in the dimer. The results again indicate that the obtained tetramer models could not represent a functionally relevant *trans*-tetrameric adhesion unit (Fig. 3).

Neither avidity nor native-like orientation can facilitate the formation of *trans*-tetramers. The bead aggregation assays (BAA) represents a simple and robust way to assess homo-oligomerization of proteins. Because of its advantages in terms of providing avidity as well as possibility to mimic natural orientation(s), this approach has been often used to characterize cell-adhesion molecules^{43–46}. We have performed BAAs with various EpCAM variants all C-terminally fused to hinge-and-Fc regions of human IgG (Fig. 4a). The purpose of the fusion is dual: (1) by locking the two EpCAM subunits in close proximity via a di-sulfide bond in the FC regions it stabilizes the *cis*-dimer and thereby ensures that only higher-order oligomerization can facilitate the aggregation of beads, and (2) binding of the Fc region to the bead-immobilized Protein A orients EpCAM in a correct, native-like orientation, where extracellular domains point outwards of the surface. The EpCAM variants used were ngEpEX and gEpEX (thereby considering also the glycosylation, albeit non-native as produced in insect cells), and gEpFL; ngEpFL was not used since the Fc-fusion could not be obtained in sufficient amounts. As a positive control an Fc fusion of extracellular domain (EC1-EC5) of human E-Cadherin was used (E-CadEX-Fc) - such fusion has been previously reported to successfully induce bead aggregation^{43,44}.

Results indicate that none of the used forms of EpCAM can induce any significant bead aggregation, while E-CadEX-Fc induced bead aggregation in presence of calcium ions, which was blocked upon treatment with EDTA (Fig. 4b). Bead aggregation ratios of EpEX fusions are not significantly different than those of Protein A beads only (Fig. 4c), suggesting EpCAM dimers attached to separate beads do not interact through their extracellular domains in a homophilic manner. Slightly increased aggregation ratio was observed in the case of gEpFL-Fc, however not statistically significant and only comparable to the ratio of E-CadEX-Fc in the presence of EDTA that is physiologically irrelevant. Considering the protein aggregation during gEpFL-Fc purification, this bead aggregation is probably mediated through TM-region embedded in micelle and is not the consequence of interactions between extracellular domains.

Aggregation of gEpEX was further analyzed at various pH values and without calcium ions to ensure all relevant physiological conditions were considered, however none of those conditions resulted in any significant bead aggregation (Fig. 4c).

Taken together, bead aggregation assay is in line with our previous observations that EpCAM *cis*-dimers do not interact in a manner that could result in formation of cell-cell contacts (i.e. *trans*-dimerization of *cis*-dimers on neighboring cells). Furthermore, even avidity, inherent by design in BAA, does not contribute to formation of such inter-bead higher-order oligomers. This indicates that neither correct native-like orientation nor high local concentration of EpCAM molecules can facilitate the formation of *trans*-tetramers.

EpCAM forms *cis*-dimers but not *trans*-oligomers *in vivo*. To analyze homophilic interactions between EpCAM molecules in their native setting, i.e. cellular environment, including native-like glycosylation, we employed FRET to probe the distances between various fluorescently-labeled EpCAM variants. In general, donor and acceptor fluorophores attached to the proteins under study have to be no more than 10 nm apart for FRET to occur. To date, several FRET-based methods have been developed to measure direct protein-protein interactions in living or fixed cells⁴⁷ and in a properly designed experiment, the presence of FRET is considered a direct indication of protein-protein interaction. For our analysis we used Fluorescence Lifetime Imaging Microscopy (FLIM-FRET), which is a very robust method because variations in excitation intensity, inner filtering, moderate donor photobleaching and detector sensitivity do not influence fluorescence lifetime⁴⁸. In FLIM-FRET, existence of FRET can be inferred from a drop of fluorescence lifetimes (τ), as the presence of acceptor fluorophore offers an additional deactivation pathway from the donor-excited state⁴⁸.

We observed lifetimes of different combinations of N- or C-terminal EpCAM fusions with an EGFP variant sfGFP or/and mCherry. EGFP variants and mCherry are well characterized FRET pairs, suitable for FLIM with $r_0 = 5.4$ nm⁴⁹. The distance between N-terminals in *cis*-dimer (~7 nm) and the maximal distance between N-terminals in a *trans*-tetramer (~5 nm, assuming a D2 symmetric *trans*-tetramer with maximal, 90° angle at y-axis) fall within the measurable FRET range for this FRET pair. Donor lifetimes were measured in cells transfected with either N-terminal donor only (GFP_{EX}) to measure τ of donor fluorophore when FRET is not possible, N-terminal donor covalently linked to N-terminal acceptor on the same protein (GFP-Cherry_{EX}), to assess the maximal drop of τ due to FRET, and N-terminal donor and C-terminal acceptor transfected in the same cell (GFP_{EX}/Cherry_{IN}) as a negative control for distances exceeding FRET range. Next, we measured donor lifetimes in N-terminal donor and N-terminal acceptor transfected in the same cell (GFP_{EX}/Cherry_{EX}) to observe intra-cellular *cis*-oligomerization, and of mixed cells transfected with either donor only or acceptor only, to form donor-acceptor pairs at the areas of cell-cell contacts (GFP_{EX}, Cherry_{EX} mixed), which would result in FRET due to inter-cellular *trans*-oligomerization (Fig. 5a). Two different EpCAM negative cell lines were used to observe potential changes in different cell types: human epithelial cell line HEK 293 T (Fig. 5a,b) and CRISPR/Cas EpCAM-knockout colorectal adenocarcinoma cell line HCT8 #L13²⁰ (Supplementary Fig. 4).

In our case, GFP_{EX} exhibited biexponential fluorescence decay with average lifetime of 2.14 ± 0.05 ns (Fig. 5c). Significant drop in the average lifetime ($\tau_{\text{avg}} = 1.2 \pm 0.1$ ns) was observed in case of covalently linked donor and acceptor (Fig. 5c). As expected, no change in donor lifetime was observed in GFP_{EX}/Cherry_{IN} ($\tau_{\text{avg}} = 2.13 \pm 0.06$ ns) (Fig. 5c), where FP are on the opposite site of the membrane, too far apart for FRET. On the other hand, GFP_{EX}/Cherry_{EX} exhibited average lifetime of 1.6 ± 0.15 ns, indicating EpCAM does form

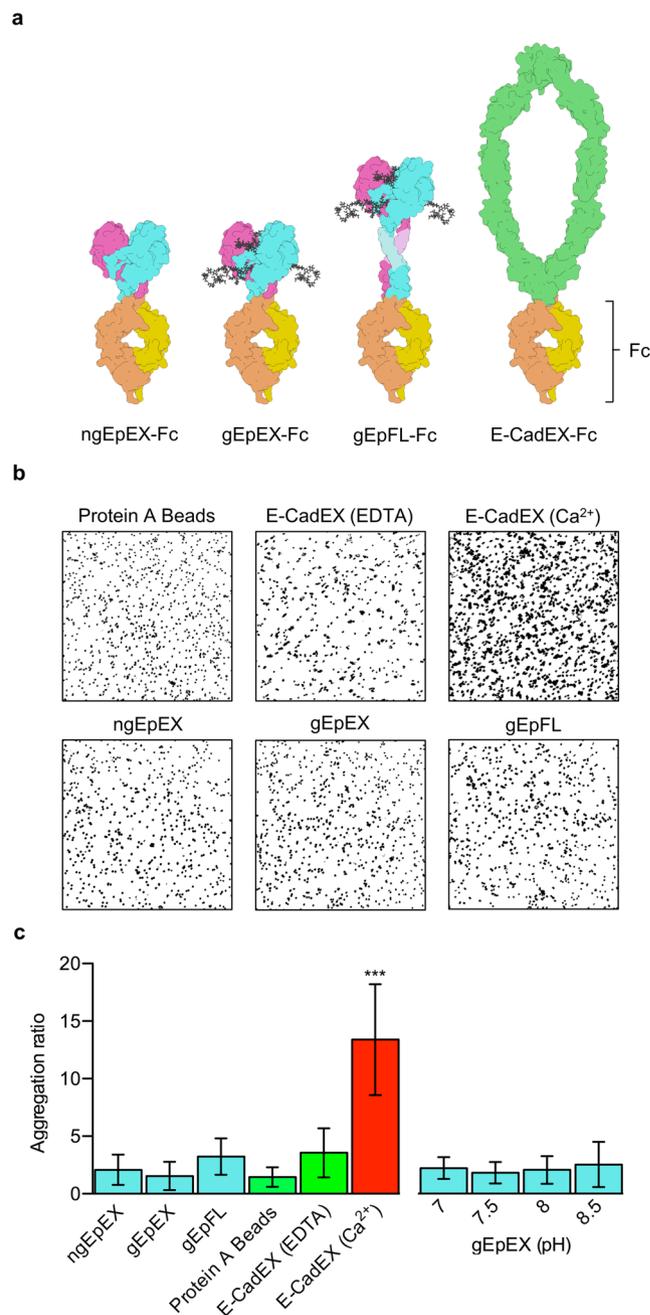


Figure 4. Bead Aggregation Assays. **(a)** Schematic representation of analysed proteins. Subunits in EpEX and EpFL are depicted as in Fig. 1. Glycosylation is depicted as grey sticks. E-CadEX dimer is green and Fc-dimer is orange and yellow. **(b)** Representative bead aggregation images after image analysis with ImageJ. Bead aggregation is seen as clustering of black spots. **(c)** Comparison of Aggregation ratios. Values represent mean values of 15 independent images with s.d., *** $p < 0.001$, one-way ANOVA test with Bonferroni post hoc analysis.

cis-oligomers *in vivo* (Fig. 5c). To answer whether the observed difference in the lifetime is a result of dimerization or even higher-order oligomerization, further experiments would be required.

When cells expressing either sfGFP-EpFL or mCherry-EpFL were mixed (GFP_{EX}, Cherry_{EX} mixed), no statistically significant change in lifetime was observed at the contact area ($\tau_{\text{avg}} = 2.14 \pm 0.06$ ns) (Fig. 5c). To ensure fixation was not the cause for the lack of observed inter-cellular FRET, this experiment was also performed in live HEK 293 T cells, yielding similar results (Supplementary Fig. 5a). Furthermore, lifetimes at cell-cell contact areas did not significantly differ from lifetimes of areas where only donor FP was present (Supplementary Fig. 5b). This provides evidence that although EpCAM may be localized at the areas of the cell-cell contact as reported previously, EpCAM molecules on the opposing cells are not sufficiently close nor involved in a direct interaction.

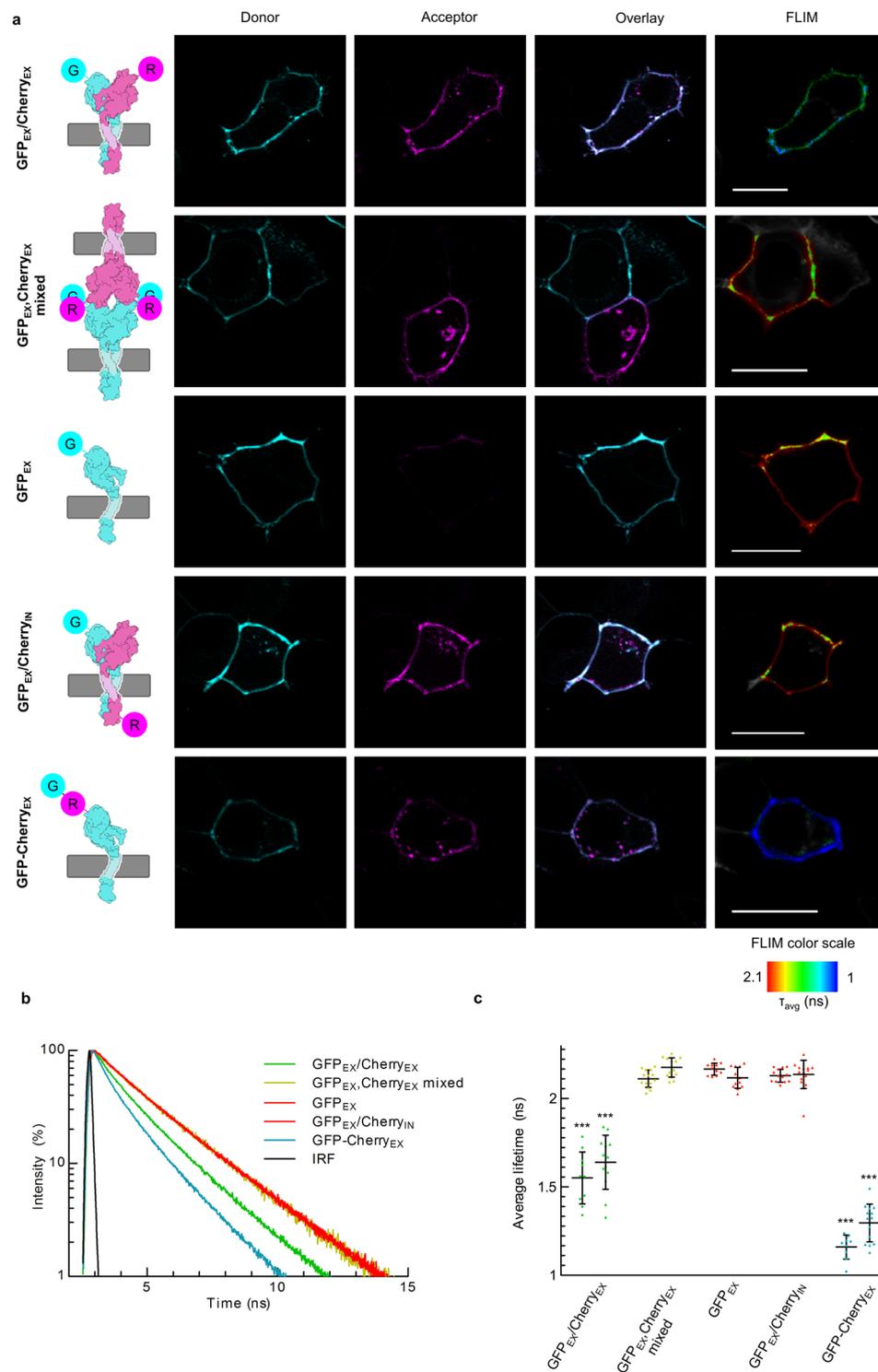


Figure 5. FLIM-FRET (HEK 293 T). **(a)** Analysed combinations of fluorescently tagged EpCAM proteins. sfGFP (G) fluorescence is colored cyan and mCherry (R) fluorescence is colored magenta. The same color scheme applies to schematic representations on the left. White line represents 20 μm . FLIM color scale is the same in all presented FLIM measurements. **(b)** Representative fluorescence lifetime decays for each analysed combination. **(c)** Mean lifetimes with s.d. of all analysed combinations. For each combination, results obtained in HEK 293 T cells are presented on the left and results obtained in HCT8^{EpCAM-} on the right, *** $p < 0.001$, one-way ANOVA test with Bonferroni post hoc analysis, compared to mean values of negative controls. Color scheme is the same as in figure b.

Discussion

Initial analysis of the function of EpCAM determined a role in cell-cell adhesion^{19,26} through homophilic interactions^{23,24}. These assigned functions led to its name as epithelial cell adhesion molecule. Ever since and despite seemingly contradicting observations regarding its role in formation of cell-cell contacts^{20,25,31}. EpCAM's ability to form adhesion units through homo-oligomerization of extracellular domains was well accepted. The results generated by our comprehensive approach using a variety of methods, constructs (full-length and truncated form of EpCAM, with/without fusion) and their variants (non-glycosylated, non-native glycosylated and native glycosylated) both under *in vitro* and *in vivo* conditions present strong evidence that such interactions are highly unlikely and that EpCAM thus cannot function as a homophilic CAM, at least not in a manner in which it was initially proposed^{23,24}.

The SAXS analysis of EpEX oligomeric state in solution clearly demonstrates that the EpEX exists predominantly in its dimeric form (Fig. 2). First, both shape and size as well as comparison of calculated SAXS profiles with experimental SAXS profiles (Fig. 2; Supplementary Fig. 1) confirm the initial assumption that the dimer observed in the crystal structure represents the actual structure of the dimer in solution²⁵. Furthermore, this provides additional support for the assumption that some of the initial studies on EpCAM oligomerization, claiming that EpEX is a monomer up to a concentration of 6 mg/ml²⁴, were premature. We can only speculate what contributed to such discrepancy in results; one possible reason is that the initial conclusion was based on sedimentation equilibrium analysis which lacks sufficient sensitivity as compared to SAXS. While there could potentially still be an equilibrium between monomeric and dimeric form even at low concentrations, EpEX clearly does not partake in formation of any higher-order oligomers consistent with its function as a homophilic CAM (Supplementary Fig. 1). Although such analyses of soluble ectodomains of adhesion proteins do not completely represent the conditions *in vivo*, the applicability of results to full-length proteins have been demonstrated in many other similar cases including Cadherins^{50–52}, PECAM-1⁵³, NCAM⁵⁴, JAM-A⁵⁵, LI⁵⁶, Nectin⁵⁷, Desmocollins and Desmogleins⁵⁸. This is not a surprise as large cell-cell interaction-mediating ectodomains of type I transmembrane adhesion proteins frequently represent the major portion of the protein. Function of the transmembrane region and the intracellular domain is, in this respect, mainly to anchor the protein to the membrane and the cytoskeleton, which was similarly claimed for EpCAM⁵⁹. Furthermore, consideration of concentrations used in our SAXS analysis with regard to K_d values of inter-cellular interactions of other adhesion proteins confirms that the concentrations were high enough to enable formation of functionally-relevant higher-order oligomers. In the light of this we did not observe any sign of oligomerization at concentrations roughly an order of magnitude higher than the K_d for oligomerization of E-Cadherin ectodomain^{50,51}, which is considered as one of the weaker cell-cell interactions.

The only method used to date to study both structure and stoichiometry of EpCAM oligomers was chemical cross-linking^{23,24}. We have included it also in our repertoire, however we have supplemented it with MS-based cross-link identification. This enabled us to match the identified cross-links to subunit and dimer structure of EpCAM as well as to the models of the proposed tetramer. As already reported in the initial studies, we have also observed bands corresponding to molecular mass of di-, tri- and tetramer after cross-linking either EpEX or EpFL (Fig. 3a). Presence of bands corresponding to tri- and tetramer is not in agreement with SAXS experiments which do not indicate higher-than-dimer oligomeric forms in solution at the same protein concentration (1 mg/ml). This calls for careful investigation whether the apparent presence of higher-order oligomers is a consequence of cross-linking artefacts or a result of biologically relevant interactions. Our structural modelling based on spatial restraints obtained from XL-MS suggest the cross-linking artefacts since the best model explaining the experimental data is based on random tetramers without the ability to support formation of the postulated cell-cell contacts (Fig. 3e). The bands corresponding to trimers were separately used for structural modelling as they can be assumed to stem from partially cross-linked tetramers. However, due to limitations of modeling of homo-oligomeric proteins based on cross-link-derived distance restraints, namely the oligomeric state ambiguity of identified connections and the possibility of false positives, a definite disproval of *trans*-tetramer existence is not possible. *Trans*-tetramers could be present in our samples however their identification was hindered due to beforementioned restrictions or the fact that the *trans*-tetramer specific cross-links simply weren't detected in MS. Still, our extensive analysis clearly demonstrates tetramer existence cannot be directly implied based solely on apparent molecular weight of cross-linked species, as it was done in previous reports^{23,24}.

To sum up, relatively high concentrations of cross-linker needed to observe significant bands corresponding to tetramer of ngEpEX and the fact that all four cross-links not found in subunit and dimer can be better explained by random interactions between two dimers indicate that the appearance of the band corresponding to higher-order oligomers is more likely an artefact than a consequence of functionally relevant *trans*-tetramers.

Similarly, no aggregation was observed in BAA using either ngEpEX-Fc or gEpEX-Fc. While dimerization via di-sulfide bonds in Fc-regions ensures EpEX subunits are covalently linked in a dimer, it could also impair proper formation of the EpEX *cis*-dimer, which is a prerequisite for aggregation via the formation of a *trans*-tetramer. However, due to the inherent flexibility of hinge regions in IgGs⁶⁰ and the stability of EpEX *cis*-dimer demonstrated in this and other research²⁵, and matching distance between N-terminals in FC-region dimer and N-terminals in EpCAM transmembrane region models, we believe it is highly unlikely proper formation of the EpEX *cis*-dimer would be completely abolished. Furthermore, such potential structural hindrance would not affect proper dimerization of gEpFL in gEpFL-FC fusion, where EpCAM's transmembrane region and EpIC are located between the EpEX and the hinge-and-FC region. Small, statistically insignificant, difference from negative control aggregation ratio could be detected using gEpFL-Fc (Fig. 4c,d). Since neither ngEpEX-Fc nor gEpEX-Fc induced any bead aggregation at all, this can't be attributed to *trans*-interactions between ectodomains. The more likely explanation is that EpFL is prone to *cis*-aggregation. This hypothesis is in agreement with previously published results where the presence of bands corresponding to tetramers in *in vivo* cross-linking experiment was

observed even when single, non-interacting cells were examined (in this case the authors concluded the adhesion unit is a *trans*-octamer, without any direct evidence of its existence)²³.

To place the system under study into cellular context we employed FLIM-FRET analysis, which did show evidence of *cis*-oligomerization (Fig. 5, Supplementary Fig. 4, experiment labeled GFP_{EX}/Cherry_{EX}). However, the experimental design prohibited us from elucidating the stoichiometry of interactions. Based on other results presented in the present and in previous publications²⁵, we speculate that the observed change in donor lifetimes was a consequence of *cis*-dimerization. On the other hand, no evidence of *trans*-oligomerization even though both the donor and the acceptor labeled EpCAM on the opposing cells were colocalized at the area of cell-cell contact (Fig. 5, Supplementary Fig. 4, experiment labeled Cherry_{EX} mixed). Combined with our *in vitro* experiments discussed above, this provides conclusive evidence that EpCAM is not able to form inter-cellular *trans* interactions, which are prerequisite for its function as a homophilic CAM.

If EpCAM indeed is a CAM, our results show that its function like one can't be directly linked to its homo-oligomerization as it has been widely accepted for more than 20 years. Even more, we believe that the most feasible description of EpCAM is that it is in fact not a cell-cell adhesion molecule at all. Our results in combination with those previously reported by Tsaktanis *et al.*²⁰ support this hypothesis both *in vitro* and *in vivo*. Furthermore, EpCAM shares no structural similarity to any other superfamily of known adhesion molecules. Also, the role in cell-cell adhesion has never been described for its closest homologue Trop2³¹ with which it shares 67% amino acid sequence similarity. It should also be noted that the Fornaro *et al.* failed to reproduce the initial results of EpCAM's direct involvement in cell-cell adhesion in transfected murine fibroblasts³¹. Similarly, no effect on cell segregation upon transfection was observed in thymic epithelial cells⁶¹. Additionally, EpCAM mediated adhesion has been described as very weak; therefore, zipper- or cluster-like extensive contacts would be expected to compensate for weak single inter-cellular interactions via avidity, however they have never been observed^{23,62}.

Our results clearly demonstrate that EpCAM partakes in *cis*-oligomerization *in vivo*, however since no other interactions between ectodomains exist both *in vitro* and *in vivo* some aspects EpCAM biology should be revisited in the future. First, the role of alleged *trans*-interactions in inducing RIP should be addressed in detail. If homophilic interactions as believed to date¹⁶, aren't the cause for its initialization, we speculate that an interaction with a yet unidentified ligand might be involved. Also, it has been already demonstrated that the dimeric form would structurally obstruct cleavage by ADAM and BACE1 during RIP²⁰. Similarly, dimerization of the substrate has also been shown to hinder processing by γ -secretase in the case C-terminal fragment of the amyloid β protein-precursor (APP CTF β)⁶³. In the light of this the role of the EpCAM *cis*-oligomerization and its regulation, particularly with regard to its interaction with other proteins, still remains to be studied in detail.

Even without direct involvement in formation of cell-cell contacts, EpCAM remains an important mediator of cell-cell adhesion, since it is involved in negative regulation of both classical E-Cadherin mediated cell-cell adhesion^{27,28} and positive regulation of tight junction formation via a direct interaction with claudin-7^{29,30}. These interactions combined with its ever more evident signaling function are still sufficient to explain its diverse role in epithelial morphogenesis⁶⁴, organization⁶⁵ and cancer¹². Therefore, we believe that properly addressing the function of EpCAM, including refuting its property as a homophilic CAM, will help in elucidating its various functions. To avoid misconception, we also support that the molecule is renamed to Epithelial Cell Activating Molecule, as previously suggested^{20,66}.

Methods

Protein cloning, expression and purification. ngEpEX and gEpEX were expressed in insect cell line *Spodoptera frugiperda* (Sf9, Thermo Scientific, USA) and purified as described before²⁵.

EpFL forms were expressed in similar manner. 3 days post transfection with baculoviral stock, cells were collected by centrifugation for 10 min at 10,000 \times g, resuspended in 20 mM HEPES, pH 7.5, 150 mM NaCl, and pelleted again. The pellet was then resuspended in 50 mM HEPES, pH 7.5, 150 mM NaCl, 20% (v/v) glycerol, 10 mM EDTA, 1 mM PMSF, and lysed using a Teflon homogenizer. Membrane fraction was collected by centrifugation for 1 h at 40,000 \times g. To solubilize EpFL, 50 μ l/10⁶ cells of extraction buffer containing 20 mM HEPES, pH 7.5, 150 mM NaCl, 20% (v/v) glycerol, 2% (v/v) Triton X-100 was added. The suspension was stirred for 1 h at 4 °C. Insoluble fraction was removed by centrifugation for 1 h at 40,000 \times g. Soluble fraction was applied to cComplete™ His-Tag Purification Column, which was previously equilibrated using 20 mM HEPES, pH 7.2, 300 mM NaCl, 15% (v/v) glycerol, 0.05% (v/v) Triton X-100 (reduced form). Elution was done using an imidazole gradient (final concentration of 500 mM). Fractions containing EpFL were dialyzed against the binding buffer overnight and then reappplied to the same column. After elution and dialysis using the same procedure as previously, samples were concentrated and purified using size-exclusion chromatography column (Superdex 200) equilibrated using 10 mM HEPES, pH 7.5, 250 mM NaCl, 0.05% (v/v) Triton X-100 (reduced form).

EpEX or EpFL fused to IgG1 heavy chain hinge and Fc region (both derived from ps521-hEpEX-Fc²⁰, a kind gift from Olivier Gires) and with added C-terminal His₆-tag were cloned into pFastBac1. Following transposition to generate recombinant bacmids in *E. coli* DH10MultiBac cells the recombinant bacmids were isolated and used to transfect *S. frugiperda* Sf9 cells. The process of baculovirus preparation and protein expression was the same as described before²⁵. For isolation, only one immobilized metal affinity chromatography and subsequently one size exclusion chromatography step was performed. Buffers were the same as the corresponding buffers used for either EpEX or EpFL.

E-CadEX-Fc fusions were designed with fusion to human E-Cadherin (aa 1–709), cloned from hE-cadherin-pcDNA3⁶⁷ (a kind gift from Barry Gumbiner, Addgene plasmid #45769), and IgG1 heavy chain hinge and Fc-region (residues 105–330). E-CadEX-Fc was cloned into pFastBacDual-Furin under polyhedrin promoter. pFastBacDual-Furin already harbored human Furin, cloned from pGEMfur⁶⁸ (a kind gift from Gary Thomas), under the control of p10 promoter. Coexpression of Furin has previously been shown to be necessary to ensure

efficient cleavage of E-Cadherin propeptide in insect cells⁶⁹. Transposition, transfection, virus amplification, protein expression and purification were done the similarly as for EpEX-Fc.

EpFL fusions to fluorescent proteins were cloned into pcDNA3.1 *myc*-His B (Life Technologies). sfGFP⁷⁰ (residues 1–238) and mCherry (residues 1–236) sequences were obtained from iGEM's Registry of Standard Biological Parts (BBa_K1365020 and BBa_K773003, respectively). N-terminal fusions were prepared in the following manner: EpCAM's native signal sequence (residues 1–23) with addition of two amino acid residues (QE) to ensure proper cleavage of signal peptide was fused to either sfGFP or mCherry, which was separated from a C-terminally located EpFL (residues 24–314) by amino acid residues GS. C-terminal fusions were comprised of EpFL with its signal peptide (residues 1–314), GS-linker and either sfGFP or mCherry. To generate GFP-Cherry_{EX}, sfGFP (1–227) was first fused to mCherry3 via the linker LESGGEDPMVSKGEE. This fusion was then cloned as other N-terminal fusions described above.

Small angle X-ray scattering (SAXS). SAXS profiles of ngEpEX were collected at concentrations ranging from 0.5 to 26.2 mg/ml at 15 °C at Stanford Synchrotron Radiation Lightsource (SSRL) Beamline BL4-2 (SLAC National Accelerator Laboratory, Menlo Park, CA). Proteins were purified as described and reference buffer matching was achieved by triple dialysis steps, each overnight at 4 °C against a buffer composed of 20 mM HEPES pH 8.0, 100 mM NaCl, 5% (v/v) glycerol. The 5% glycerol was used for radiation protection of the proteins⁷¹. All the suspensions were filtered through membranes with 0.1 μm pore size (Millipore, Bedford, MA).

The beam energy and current were 11 keV and 500 mA, respectively. A silver behenate sample was used to calibrate the q-range and detector distance. Data collection was controlled with Blu-Ice⁷². We used an automatic sample delivery system equipped with a 1.5 mm-diameter thin-wall quartz capillary within which a sample aliquot oscillated in the X-ray beam to minimize radiation damage⁷³. The sample was placed at 1.7 meter from a MX225-HE (Rayonix, USA) CCD detector with a binned pixel size of 293 μm by 293 μm. Up to twenty 1-second exposures were made for each protein sample and buffer maintained at 15 °C. Each of the diffraction images was scaled using the transmitted beam intensity, azimuthally integrated by SasTool (<http://ssrl.slac.stanford.edu/~saxs/analysis/sastool.htm>, formerly called MarParse), and averaged to obtain fully processed data in the form of intensity versus q [$q = 4\pi\sin(\theta)/\lambda$, θ = one-half of the scattering angle; λ = X-ray wavelength]. The buffer profile was subtracted from a protein profile using SasTool. The buffer subtracted SAXS profiles were initially analysed using the ATSAS package⁷⁴ to calculate radius of gyration (R_g) and maximum particle size (D_{max} ; Supplementary Table 1). Subsequently, the molecular weight of ngEpEX was estimated at multiple concentrations using SAXS MOW⁷⁵. The mean of the smaller scattering angle regions ($q < 0.15 \text{ \AA}^{-1}$) of the lower concentration profiles (0.5–1.5 mg/ml) and the mean of the wider scattering angle regions ($q > 0.12 \text{ \AA}^{-1}$) of the higher concentration profiles (2.0 to 26.2 mg/ml) were merged to obtain the final experimental SAXS profiles. The *ab initio* shape reconstruction for each sample (Fig. 2C) was generated from the corresponding SAXS profile by running DAMMIF⁷⁶ 20 times and then refined through an additional 40 DAMMIN⁷⁷ runs followed by superposition and averaging with DAMAVER⁷⁸.

Molecular envelope was generated using merged SAXS profiles by running DAMMIF⁷⁶ 15 times followed by superposition and averaging with DAMAVER⁷⁸. As EpEX dimer is symmetric, P2 symmetry was assumed however the anisometry level was set to unknown. Final representation of molecular envelope and fitting of EpEX dimer structure was done using UCSF Chimera^{79,80}.

Oligomeric state modeling based on SAXS. Oligomeric state analysis was performed *via* comparison of calculated SAXS profiles to experimental SAXS profile using FoXS^{33,34} and Multi-FoXS³⁴, plus by fitting of structure models into the *ab initio* shapes calculated using DAMMIF⁷⁶. Our initial analysis (Fig. 2A,B) showed that calculated R_g and MW values very well correspond to a dimer even at the lowest concentration.

First, a complete model of EpEX dimer was constructed using the EpEX crystal structure (PDB: 4MZV; residues 24–258 were used). C-termini of EpEX dimer corresponding to residue stretch SMQGLK (residues 259–265) along with LE from translated restriction site and His₆-tag were modeled with MODELER⁸¹. 30,000 initial models were generated and only top-scoring 10,000 models (according to SOAP score⁸²) were used in further analysis. SAXS profiles calculated from dimer structures using FoXS were fitted to merged experimental data in order to get the best fitting model to be used as a building block for tetramer modeling.

Next, tetramer models were generated by combining two dimer structures using PatchDock^{36,37}. High accuracy sampling with final clustering set to RMSD 2 Å yielded 44,098 tetramer models with random dimer-dimer orientations. Again, FoXS was employed to calculate their SAXS profiles.

Finally, multi-state modeling with a combined dataset of SAXS profiles calculated from 10,000 dimer and 44,098 tetramer models was done using Multi-FoXS in order to see which combination of the models best describe the experimentally determined SAXS profile. To ensure all possible combinations are considered, χ value percentage threshold for profile similarity was set to 0.1, minimal weight threshold for a profile to contribute to the ensemble was set to 0.01, and the number of combinations expanded to the higher state was set to 100,000. Although the maximal number of states was set to four, fitting stopped at two states, as inclusion of an additional state would fail to significantly improve the fit.

Chemical cross-linking coupled with mass spectrometry. Purified ngEpEX and ngEpFL protein (equivalent to 700 pmol) was mixed with 5 ×, 25 ×, or 125 × molar excess of disuccinimidyl suberate (DSS) cross-linker (ThermoFisher Scientific) in a final volume of 23 μl of 20 mM HEPES pH 8.0, 100 mM NaCl. The reaction was carried out at 37 °C at 1000 RPM on a thermomixer and stopped after 30 min by adding 1 μl of 1 M Tris pH 8.0 and 8 μl of 4 × SDS-PAGE reducing loading buffer (250 mM Tris pH 6.8, 8% SDS, 300 mM DTT, 30% glycerol, 0.02% bromophenol blue). Cross-linked proteins were separated by SDS-PAGE on a 4–20% TGX gel (Criterion) and stained with AcquaStain Protein Stain (Bulldog Bio). Gel bands were excised and diced

into 1 mm × 1 mm cubes. Gel pieces were dehydrated three times by adding enough extraction buffer (25 mM NH₄HCO₃, 50% ACN) to cover pieces, vortexing for 10 min, and discarding supernatant. Residual liquid was completely dehydrated by SpeedVac. Samples were then reduced and alkylated by first rehydrating gel pieces in enough 15 mM TCEP, 25 mM NH₄HCO₃ to cover pieces. Reduction reaction proceeded at room temperature for 20 min and then fresh 0.5 M Iodoacetamide, 25 mM NH₄HCO₃ was added to a final concentration of 50 mM. Reaction proceeded in the dark at room temperature for 20 min. Supernatant was removed and gel pieces were washed once with 25 mM NH₄HCO₃, then dehydrated with extraction buffer 2 times (vortexing for 5 min). Residual liquid was removed by SpeedVac. Gel pieces were rehydrated in trypsin buffer (0.5 ng/μl mass spec grad trypsin in 25 mM NH₄HCO₃) with enough buffer to cover pieces (typically 10–50 ng trypsin per gel slice). Digest reaction proceeded at 37 °C overnight. Supernatant with resulting peptides was collected, and residual peptides in gel bands were extracted 2 times with 50% ACN, 5% Formic Acid (vortexing for 10 min). A final extraction with 20 μl 100% ACN (vortex 5 min) was collected and all extractions combined to a single extraction tube. Peptides were dried by SpeedVac and kept frozen until LC-MS/MS analysis.

Digested peptides were analyzed by LC-MS/MS using a Thermo Fisher Scientific Easy-nLC 1000 coupled to a dual-pressure linear ion trap (Velos Pro) Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, San Jose, CA). LC-MS/MS was carried out as follows. Prior to LC-MS/MS analysis, peptides were dissolved in 12 μl 3% ACN, 2% Formic Acid. For each sample, 2 μl were loaded onto an online 75 μm × 30 cm fused silica IntegraFrit capillary packed with 1.9 μm Reprosil-Pur C18 AQ reversed-phase resin (Dr. Maisch-GmbH). Peptides eluted by the following gradient program: 5% to 35% ACN, 0.1% formic acid in 10 min, 35% to 95% ACN, 0.1% formic acid in 5 min, and 95% ACN, 0.1% formic acid for 10 min (delivered at a flowrate of 300 nl/min). For each MS cycle, one full MS scan (150–1500 m/z, resolution of 120,000) in the Orbitrap was followed by 20 data-dependent MS/MS scans targeting the twenty most intense ions (with charge state exclusion of all 1+, 2+, and 3+ ions). Selected ions were fragmented by normalized collision energy (setting of 35%) and acquired in the linear ion trap, with any previously acquired ion being dynamically excluded for 20 seconds. Monoisotopic masses of parent ions and corresponding fragment ions, parent ion charge states and ion intensities from LC-MS/MS spectra were extracted using in-house software based on Raw_Extract script from Xcalibur v2.4. Data was searched against a database with the sequences of the EpCAM truncated His-tagged variants, and cross-linked peptides were identified using the Cross-linked Peptide Analysis feature of the BatchTagWeb application on the Protein Prospector search engine (v5.14.2)³⁹ DSS was selected as the cross-linker, 20 ppm and 0.5 Da were set as the respective parent and fragment mass tolerances. Default parameters were selected according to the tutorial provided by Protein Prospector (<https://vimeo.com/97952461>).

Oligomeric state modeling based on XL-MS. SASD were calculated using Jwalk (version 1.0) algorithm⁴⁰. Two minor modifications were included in the process. First, cross-links with Euclidean distances longer than the cutoff (33 Å) were excluded from SASD calculation to reduce the overall time needed for the analysis. Second, algorithm was modified to account for all possible SASDs of each cross-link, i.e. intra- and inter-subunit.

To each calculated SASD MNXL score was assigned according to published parameters⁴⁰:

$$\text{MNXL score [SASD]} = \begin{cases} N(18.62, 35.94) & \text{if SASD} \leq 33 \\ -0, 1 & \text{else} \end{cases}$$

To avoid overscoring of the symmetric models only the best MNXL score was considered, even if there were multiple SASDs calculated for a given cross-link. When scoring tetramer models, the negative penalty for SASD > 33 Å was omitted if the same cross-link had a positive MNXL score in a dimer. The reason for this was to favour the tetramer models matching those cross-links which were non-accessible in the dimer models.

Using this procedure, we scored SASDs using the following models: a monomer model (generated as a subunit of a dimer structure D-21988, described previously), a dimer structure (D-21988), all 44098 tetramer models, 12 D2 symmetric *trans*-tetramer models (generated with Symmdock^{37,83}) and 1000 models extracted from corresponding 1000 frames in 10 ns (one frame per 10 ps) MD simulation of monomer a model to account for its flexibility. For simulations NAMD 2.11⁸⁴ with CHARMM-27 force-field was used. Briefly, the monomer model with C-terminus and His₆-tag modelled in an extended conformation was solvated using a water cube (20 Å margin on each side) and charge-neutralized (addition of Na⁺/Cl⁻ ions). Following minimization for 1000 steps the system was equilibrated for 10 ns (2 fs/step) at 310 K using constant temperature Langevin dynamics, constant pressure via Noose-Hoover Langevin piston and Particle Mesh Edward for full-system periodic electrostatics. Model structures were extracted from the trajectory using VMD⁸⁵.

Additionally, we generated an average occupancy grid for a tetramer model based on Euclidean distances using DISVIS^{41,42}. As DISVIS doesn't address the problem of chain ambiguity which is present in homo-oligomeric cross-links, chain IDs to cross-links non-accessible in the dimer were assigned in a way that the resulting cross-links had the shortest possible distance in the predicted *trans*-tetramer model. Furthermore, symmetry was imposed by adding symmetric cross-links, e.g. for each chain A to chain D cross-link we also added a chain C to chain B cross-link. Average occupancy grid of models which satisfied the highest number of restrains was generated with UCSF Chimera^{79,80}.

Bead aggregation assays. Bead aggregation assays were performed using Fc-chimeric fusion proteins immobilized by the native strong interaction to Protein A-covered magnetic beads. These beads were prepared by coupling Protein A (Sigma Aldrich) to magnetic beads with a diameter of 1 μm (Dynabeads[®] MyOne[™] Carboxylic Acid, Thermo Fisher Scientific) according to manufacturer's instructions.

Fc-chimeric proteins were incubated together with Protein A-beads for 20 minutes at room temperature, washed twice with PBS buffer containing 0.1% Tween-20, and resuspended in an appropriate buffer. For resuspension of ngEpEX-Fc, gEpEX-Fc and E-CadEX-Fc(Ca²⁺) 20 mM HEPES, pH 8.0, 100 mM NaCl, 2 mM CaCl₂, 0.1% (v/v) Tween-20 was used. In the case of gEpFL-Fc Tween-20 was replaced by Triton X-100. E-CadEX-Fc(EDTA), was resuspended in the same buffer containing 2 mM EDTA instead of CaCl₂. Similarly, for evaluating the possible influence of pH on gEpEX-Fc-induced aggregation, a buffer with the same composition without either CaCl₂ or EDTA was used and the pH value accordingly adjusted; for pH 8.5, HEPES was replaced with Tris-HCl.

Following resuspension beads were incubated with rotation for 30 minutes at room temperature and then transferred to a glass-bottomed slide (μ -Slide 8 Well Glass Bottom, Ibidi) where they were allowed to settle for 30 minutes. Differential interference contrast (DIC) images were captured using Confocal laser scanning microscope Leica TCS SP8 at 600 \times magnification. For each sample, five images were used in analysis. Experiments were done in triplicates, starting with the preparation of Protein A-coupled magnetic beads, resulting in total of 15 images per sample.

Image analysis was done using ImageJ⁸⁶. To extract particle sizes the following set of commands was applied: Smooth, Find Edges, set AutoThreshold, Convert to Mask, Fill Holes, and Analyze particles. In each image, 10 single bead particles were manually selected and their average area and standard deviation was calculated. Particles with an area larger than the average plus twice the standard deviation were considered as aggregates⁴⁵. Furthermore, particles with an area smaller than the average minus twice the standard deviation were considered as artefacts and were excluded from the analysis. Bead aggregation ratios were calculated as a ratio between the total area of aggregates divided by the total area of single beads⁴⁵.

Transfection and treatment of cells for FLIM-FRET. FLIM-FRET experiments were performed using transfected human embryonic kidney cell line HEK 293 T (Thermo Fisher Scientific) and human colon carcinoma cell line HCT8 #L13 (ref.²⁰; a kind gift by Olivier Gires) which both lack endogenous EpCAM expression. Cells were cultured in DMEM supplemented with 10% FBS and 1% Pen/Strep. Transfections were performed using Turbofect (Thermo Fisher Scientific) according to manufacturer's instructions. In case of GFP_{EX}/Cherry_{EX} and GFP_{EX}/Cherry_{IN} equimolar amounts of each plasmid was used for transfection. 24 hours after transfection, cells were trypsinized and seeded into wells of μ -Slide 8-Well ibiTreat slides (Ibidi). For FLIM-FRET analysis equal amounts of transfected cells expressing either sfGFP-EpFL or mCherry-EPFL (GFP_{EX}, Cherry_{EX} mixed) were thoroughly mixed and seeded into the same well. Cells were then left to grow for 24 hours to reach confluency before FLIM-FRET analysis to ensure formation of cell-cell contacts.

FLIM-FRET analysis. Cells were fixed with paraformaldehyde solution for 10 minutes. FLIM images of cells were accumulated for 120 s using the SymPhoTime software, which uses TCSPC (time-correlated single-photon counting) technique at 20 MHz. The donor fluorescent protein sfGFP was excited a 488 nm (laser power 2–5%) and fluorescence was detected by HyD SMD detector in range between 500–550 nm. Regions of interest were manually selected to exclude the contribution of internalized proteins. In all cases, except GFP_{EX}, Cherry_{EX} mixed, whole cell membrane region was selected. In case of GFP_{EX}, Cherry_{EX} mixed, we specifically selected the areas of cell-cell contact, where sfGFP and mCherry signals were colocalized. We achieved the best fitting results by assuming a biexponential decay. In combinations where we observed a significant decrease of average fluorescence lifetime, a triexponential decay was also tested, but the results were not statistically different. Donor lifetime in each combination was determined as amplitude weighted average lifetimes, calculated from the decay fit.

Experiments with all combinations were repeated at least two times, resulting in over 10 cells analyzed in each combination per cell line.

For time-resolved luminescence imaging, the Leica SP8 SMD (Leica Microsystems, Mannheim, Germany) confocal fluorescence microscope with TCSPC module PicoHarp 300 time-resolved unit (PicoQuant, Berlin, Germany) was used consisting of an inverted microscope (Leica Microsystems) equipped with a HCX Plane-Apochromat 63 \times / water-immersion CORR CS2 objective with NA 1.2 (Leica, Wetzlar, Germany). A pulsed WLL laser operating at 488 nm was used as an excitation source. The repetition rate was adjusted to 20 kHz. The fluorescence emission (500–550 nm) was guided through a 1 μ m pinhole and was detected with a HyD SMD. Time-resolved luminescence recordings were performed in the time-correlated single-photon counting (TCSPC) mode using a PicoHarp 300 (PicoQuant, Berlin, Germany). In all experiments, the laser power was adjusted to achieve average photon counting rates $\leq 10^5$ photons/s and peak rates close to 10^6 photons/s when recording images, thus significantly below the maximum counting rates allowed by TCSPC electronics in order to avoid pile up effects. Data acquisition and analysis were performed by the SymPhoTime64 software (PicoQuant, Berlin, Germany). Thereby, all photons collected in the full frame image were used to form a global histogram for luminescence decay fitting. For deconvolution fitting, the instrument response function (IRF) was measured daily by recording the backscattered excitation light of quenched erythrosine B.

The images were processed with LAS AF software (Leica Microsystems), SymPhoTime SPT software (PicoQuant) and ImageJ⁸⁶. Microscopy images were despeckled, background was subtracted and intensity was adjusted for better visibility. FLIM images were just despeckled and no other manipulations were performed.

Statistics. Results of BAA and FLIM-FRET experiments are represented as a mean value \pm s.d. One-way ANOVA ($P < 0.001$) with Bonferroni multiple comparison test was conducted to identify statistically significant differences in mean values.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

- Koprowski, H. *et al.* Colorectal carcinoma antigens detected by hybridoma antibodies. *Somatic Cell Genet.* **5**, 957–971 (1979).
- Herlyn, D., Herlyn, M., Steplewski, Z. & Koprowski, H. Monoclonal antibodies in cell-mediated cytotoxicity against human melanoma and colorectal carcinoma. *Eur. J. Immunol.* **9**, 657–659 (1979).
- Göttlinger, H. G., Funke, I., Johnson, J. P., Gokel, J. M. & Riethmüller, G. The epithelial cell surface antigen 17-1A, a target for antibody-mediated tumor therapy: its biochemical nature, tissue distribution and recognition by different monoclonal antibodies. *Int. J. Cancer.* **38**, 47–53 (1986).
- Simon, M., Stefan, N., Plückthun, A. & Zangemeister-Wittke, U. Epithelial cell adhesion molecule-targeted drug delivery for cancer therapy. *Expert Opin. Drug Deliv.* **10**, 451–68 (2013).
- Spizzo, G. *et al.* Prognostic significance of Ep-CAM AND Her-2/neu overexpression in invasive breast cancer. *Int. J. Cancer* **98**, 883–888 (2002).
- Spizzo, G. *et al.* High Ep-CAM expression is associated with poor prognosis in node-positive breast cancer. *Breast Cancer Res. Treat.* **86**, 207–213 (2004).
- Varga, M. *et al.* Overexpression of epithelial cell adhesion molecule antigen in gallbladder carcinoma is an independent marker for poor survival. *Clin. Cancer Res.* **10**, 3131–6 (2004).
- Fong, D. *et al.* Ep-CAM expression in pancreatic and ampullary carcinomas: frequency and prognostic relevance. *J. Clin. Pathol.* **61**, 31–35 (2007).
- Brunner, A. *et al.* EpCAM overexpression is associated with high-grade urothelial carcinoma in the renal pelvis. *Anticancer Res.* **28**, 125–8 (2008).
- Scheunemann, P. *et al.* Occult tumor cells in lymph nodes as a predictor for tumor relapse in pancreatic adenocarcinoma. *Langenbeck's Arch. Surg.* **393**, 359–65 (2008).
- Chen, X. *et al.* The overexpression of Epithelial cell adhesion molecule (EpCAM) in glioma. *J. Neurooncol.* <https://doi.org/10.1007/s11060-014-1459-5> (2014).
- Trzpis, M., McLaughlin, P. M. J., de Leij, L. M. F. H. & Harmsen, M. C. Epithelial cell adhesion molecule: more than a carcinoma marker and adhesion molecule. *Am. J. Pathol.* **171**, 386–95 (2007).
- Sundberg, M. *et al.* CD marker expression profiles of human embryonic stem cells and their neural derivatives, determined using flow-cytometric analysis, reveal a novel CD marker for exclusion of pluripotent stem cells. *Stem Cell Res.* **2**, 113–124 (2009).
- González, B., Denzel, S., Mack, B., Conrad, M. & Gires, O. EpCAM is involved in maintenance of the murine embryonic stem cell phenotype. *Stem Cells* **27**, 1782–91 (2009).
- Ng, V. Y., Ang, S. N., Chan, J. X. & Choo, A. B. H. Characterization of epithelial cell adhesion molecule as a surface marker on undifferentiated human embryonic stem cells. *Stem Cells* **28**, 29–35 (2010).
- Maetzel, D. *et al.* Nuclear signalling by tumour-associated antigen EpCAM. *Nat. Cell Biol.* **11**, 162–71 (2009).
- Maghzal, N., Vogt, E., Reintsch, W., Fraser, J. S. & Fagotto, F. The tumor-associated EpCAM regulates morphogenetic movements through intracellular signaling. *J. Cell Biol.* **191**, 645–59 (2010).
- Gerlach, J. C., Foka, H. G., Thompson, R. L., Gridelli, B. & Schmelzer, E. Epithelial Cell Adhesion Molecule Fragments and Signaling in Primary Human Liver Cells. *J. Cell. Physiol.* 1–42, <https://doi.org/10.1002/jcp.26286> (2017).
- Litvinov, S. V., Velders, M. P., Bakker, Ha, Fleuren, G. J. & Warnaar, S. O. Ep-CAM: a human epithelial antigen is a homophilic cell-cell adhesion molecule. *J. Cell Biol.* **125**, 437–46 (1994).
- Tsaktanis, T. *et al.* Cleavage and Cell Adhesion Properties of Human Epithelial Cell Adhesion Molecule (HEPCAM). *J. Biol. Chem.* **290**, 24574–24591 (2015).
- Chaves-Pérez, a *et al.* EpCAM regulates cell cycle progression via control of cyclin D1 expression. *Oncogene* **32**, 641–50 (2013).
- Denzel, S. *et al.* Initial activation of EpCAM cleavage via cell-to-cell contact. *BMC Cancer* **9**, 402 (2009).
- Balzar, M. *et al.* Epidermal growth factor-like repeats mediate lateral and reciprocal interactions of Ep-CAM molecules in homophilic adhesions. *Mol. Cell. Biol.* **21**, 2570–2580 (2001).
- Trebak, M. *et al.* Oligomeric state of the colon carcinoma-associated glycoprotein GA733-2 (Ep-CAM/EGP40) and its role in GA733-mediated homotypic cell-cell adhesion. *J. Biol. Chem.* **276**, 2299–309 (2001).
- Pavšič, M., Gunčar, G., Djinović-Carugo, K. & Lenarčič, B. Crystal structure and its bearing towards an understanding of key biological functions of EpCAM. *Nat. Commun.* **5**, 4764 (2014).
- Litvinov, S. V., Bakker, Ha, Gourevitch, M. M., Velders, M. P. & Warnaar, S. O. Evidence for a role of the epithelial glycoprotein 40 (Ep-CAM) in epithelial cell-cell adhesion. *Cell Adhes. Commun.* **2**, 417–28 (1994).
- Litvinov, S. V. *et al.* Epithelial cell adhesion molecule (Ep-CAM) modulates cell-cell interactions mediated by classic cadherins. *J. Cell Biol.* **139**, 1337–48 (1997).
- Winter, M. J. *et al.* Expression of Ep-CAM shifts the state of cadherin-mediated adhesions from strong to weak. *Exp. Cell Res.* **285**, 50–58 (2003).
- Ladwein, M. *et al.* The cell-cell adhesion molecule EpCAM interacts directly with the tight junction protein claudin-7. *Exp. Cell Res.* **309**, 345–57 (2005).
- Wu, C.-J., Mannan, P., Lu, M. & Udey, M. C. Epithelial cell adhesion molecule (EpCAM) regulates claudin dynamics and tight junctions. *J. Biol. Chem.* **288**, 12253–68 (2013).
- Fornaro, M. *et al.* Cloning of the gene encoding Trop-2, a cell-surface glycoprotein expressed by human carcinomas. *Int J Cancer* **62**, 610–618 (1995).
- Rambo, R. P. & Tainer, J. a. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* **496**, 477–81 (2013).
- Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys. J.* **105**, 962–974 (2013).
- Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. 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 (2016).
- Šali, A. & Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **234**, 779–815 (1993).
- Duhovny, D., Nussinov, R. & Wolfson, H. J. In *Algorithms in Bioinformatics* 185–200, https://doi.org/10.1007/3-540-45784-4_14 (2002).
- Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. & Wolfson, H. J. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* **33**, W363–W367 (2005).
- Petsch, S. *et al.* Concentrations of EpCAM ectodomain as found in sera of cancer patients do not significantly impact redirected lysis and T-cell activation by EpCAM/CD3-bispecific BiTE antibody MT110. *MAbs* **3**, 31–37 (2011).
- Trnka, M. J., Baker, P. R., Robinson, P. J. J., Burlingame, A. L. & Chalkley, R. J. Matching Cross-linked Peptide Spectra: Only as Good as the Worse Identification. *Mol. Cell. Proteomics* **13**, 420–434 (2014).
- Bullock, J. M. A., Schwab, J., Thalassinou, K. & Topf, M. The importance of non-accessible crosslinks and solvent accessible surface distance in modelling proteins with restraints from crosslinking mass spectrometry. *Mol. Cell. Proteomics* **12–15**, <https://doi.org/10.1074/mcp.M116.058560> (2016).
- van Zundert, G. C. P. *et al.* The DisVis and PowerFit Web Servers: Explorative and Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **429**, 399–407 (2017).

42. Van Zundert, G. C. P. & Bonvin, A. M. J. J. DisVis: Quantifying and visualizing accessible interaction space of distance-restrained biomolecular complexes. *Bioinformatics* **31**, 3222–3224 (2015).
43. Chappuis-Flament, S., Wong, E., Hicks, L. D., Kay, C. M. & Gumbiner, B. M. Multiple cadherin extracellular repeats mediate homophilic binding and adhesion. *J. Cell Biol.* **154**, 231–243 (2001).
44. Zhang, Y., Sivasankar, S., Nelson, W. J. & Chu, S. Resolving cadherin interactions and binding cooperativity at the single-molecule level. *Proc. Natl. Acad. Sci. USA* **106**, 109–114 (2009).
45. Chavda, A. P., Prole, D. L. & Taylor, C. W. A bead aggregation assay for detection of low-affinity protein-protein interactions reveals interactions between N-terminal domains of inositol 1,4,5-trisphosphate receptors. *PLoS One* **8**, e60609 (2013).
46. Emond, M. R. & Jontes, J. D. Bead aggregation assays for the characterization of putative cell adhesion molecules. *J. Vis. Exp.* e51762, <https://doi.org/10.3791/51762> (2014).
47. Bajar, B., Wang, E., Zhang, S., Lin, M. & Chu, J. A Guide to Fluorescent Protein FRET Pairs. *Sensors* **16**, 1488 (2016).
48. Pietraszewska-Bogiel, A. & Gadella, T. W. J. FRET microscopy: from principle to routine technology in cell biology. *J. Microsc.* **241**, 111–118 (2011).
49. Lam, A., St-Pierre, F., Gong, Y. & Marshall, J. Improving FRET dynamic range with bright green and red fluorescent proteins. *Nat.* **9** (2012).
50. Katsamba, P. *et al.* Linking molecular affinity and cellular specificity in cadherin-mediated adhesion. *Proc. Natl. Acad. Sci. USA* **106**, 11594–9 (2009).
51. Harrison, O. J. *et al.* Two-step adhesive binding by classical cadherins. *Nat. Struct. Mol. Biol.* **17**, 348–357 (2010).
52. Brasch, J. *et al.* Structure and binding mechanism of vascular endothelial cadherin: A divergent classical cadherin. *J. Mol. Biol.* **408**, 57–73 (2011).
53. Newton, J. P., Hunter, A. P., Simmons, D. L., Buckley, C. D. & Harvey, D. J. CD31 (PECAM-1) Exists as a Dimer and Is Heavily N-Glycosylated. *Biochem. Biophys. Res. Commun.* **261**, 283–291 (1999).
54. Johnson, C. P., Fujimoto, I., Perrin-Tricaud, C., Rutishauser, U. & Leckband, D. Mechanism of homophilic adhesion by the neural cell adhesion molecule: use of multiple domains and flexibility. *Proc. Natl. Acad. Sci. USA* **101**, 6963–8 (2004).
55. Kirchner, E., Guglielmi, K. M., Strauss, H. M., Dermody, T. S. & Stehle, T. Structure of Reovirus $\sigma 1$ in Complex with Its Receptor Junctional Adhesion Molecule-A. *PLoS Pathog.* **4**, e1000235 (2008).
56. Gouveia, R. M., Gomes, C. M., Sousa, M., Alves, P. M. & Costa, J. Kinetic Analysis of L1 Homophilic Interaction. *J. Biol. Chem.* **283**, 28038–28047 (2008).
57. Harrison, O. J. *et al.* Nectin ectodomain structures reveal a canonical adhesive interface. *Nat. Struct. Mol. Biol.* **19**, 906–15 (2012).
58. Harrison, O. J. *et al.* Structural basis of adhesive binding by desmogleins and desmocollins. *Proc. Natl. Acad. Sci.* **113**, 7160–7165 (2016).
59. Balzar, M. *et al.* Cytoplasmic tail regulates the intercellular adhesion function of the epithelial cell adhesion molecule. *Mol. Cell. Biol.* **18**, 4833–43 (1998).
60. Sandin, S., Öfverstedt, L.-G., Wikström, A.-C., Wrangé, Ö. & Skoglund, U. Structure and Flexibility of Individual Immunoglobulin G Molecules in Solution. *Structure* **12**, 409–415 (2004).
61. Guillemot, J. C. *et al.* Ep-CAM transfection in thymic epithelial cell lines triggers the formation of dynamic actin-rich protrusions involved in the organization of epithelial cell layers. *Histochem. Cell Biol.* **116**, 371–8 (2001).
62. Balzar, M. *et al.* The structural analysis of adhesions mediated by Ep-CAM. *Exp. Cell Res.* **246**, 108–21 (1999).
63. Fernandez, M. A. *et al.* Transmembrane Substrate Determinants for γ -Secretase Processing of APP CTF β . *Biochemistry* **55**, 5675–5688 (2016).
64. Trzpis, M., Bremer, E., McLaughlin, P. M. J., de Leij, L. F. M. H. & Harmsen, M. C. EpCAM in morphogenesis. *Front. Biosci.* **13**, 5050–5 (2008).
65. Salomon, J. *et al.* Contractile forces at tricellular contacts modulate epithelial organization and monolayer integrity. *Nat. Commun.* **8**, 13998 (2017).
66. Baeuerle, P. A. & Gires, O. EpCAM (CD326) finding its role in cancer. *Br. J. Cancer* **96**, 417–23 (2007).
67. Gottardi, C. J., Wong, E. & Gumbiner, B. M. E-Cadherin Suppresses Cellular Transformation by Inhibiting β -Catenin Signaling in an Adhesion-Independent Manner. *J. Cell Biol.* **153**, 1049–1060 (2001).
68. Leducs, R., Molloy, S. S., Thornell, B. A. & Thomas, G. Activation of Human Furin Precursor Processing Endoprotease Occurs by an Intramolecular Autoproteolytic Cleavage. 14304–14308 (1992).
69. Posthaus, H. *et al.* Proprotein cleavage of E-cadherin by furin in baculovirus over-expression system: Potential role of other convertases in mammalian cells. *FEBS Lett.* **438**, 306–310 (1998).
70. Pédélecq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24**, 79–88 (2006).
71. Kuwamoto, S., Akiyama, S. & Fujisawa, T. Radiation damage to a protein solution, detected by synchrotron X-ray small-angle scattering: dose-related considerations and suppression by cryoprotectants. *J. Synchrotron Radiat.* **11**, 462–468 (2004).
72. McPhillips, T. M. *et al.* Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines. *J. Synchrotron Radiat.* **9**, 401–406 (2002).
73. Martel, A., Liu, P., Weiss, T. M., Niebuhr, M. & Tsuruta, H. An integrated high-throughput data acquisition system for biological solution X-ray scattering studies. *J. Synchrotron Radiat.* **19**, 431–434 (2012).
74. Franke, D. *et al.* ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50**, 1212–1225 (2017).
75. Fischer, H., de Oliveira Neto, M., Napolitano, H. B., Polikarpov, I. & Craievich, A. F. Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *J. Appl. Crystallogr.* **43**, 101–109 (2010).
76. Franke, D. & Svergun, D. I. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346 (2009).
77. Svergun, D. I. Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing. *Biophys. J.* **76**, 2879–2886 (1999).
78. Volkov, V. V. & Svergun, D. I. Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36**, 860–864 (2003).
79. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
80. Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. *J. Struct. Biol.* **157**, 281–287 (2007).
81. Webb, B. & Sali, A. In *Current Protocols in Bioinformatics* **2014**, 5.6.1–5.6.32 (John Wiley & Sons, Inc., 2014).
82. Dong, G. Q., Fan, H., Schneidman-Duhovny, D., Webb, B. & Sali, A. Optimized atomic statistical potentials: assessment of protein interfaces and loops. *Bioinformatics* **29**, 3158–3166 (2013).
83. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. & Wolfson, H. J. Geometry-based flexible and symmetric protein docking. *Proteins Struct. Funct. Bioinforma.* **60**, 224–231 (2005).
84. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
85. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
86. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).

Acknowledgements

We thank D. Schneidman-Duhovny for the help with molecular modelling, and T. Matsui and T.M. Weiss at SSRL, SLAC National Accelerator Laboratory, for assistance with collecting SAXS data. This work was supported by the Slovenian Research Agency (grants J1-7119, PS-0140, PS-0207).

Author Contributions

A.G., M.P. and B.L. designed the project. A.G. prepared the samples for all experiments, performed structural modelling and did the final data analysis with the help of M.P.; S.J.K. collected and analysed SAXS data. R.M.K. prepared, recorded and analysed CX-MS data. A.G. performed and analysed BAA. M.B. helped with obtaining and interpretation of FLIM-FRET data. A.G., M.P. and B.L. prepared the manuscript with the assistance from S.J.K., R.M.K., M.B., N.K. and A.S.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-31482-7>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018