Extracellular Domains I and II of cell-surface glycoprotein CD44 mediate its trans-homophilic dimerization and tumor cluster aggregation

Received for publication, July 18, 2019, and in revised form, January 16, 2020. Published, Papers in Press, January 22, 2020, DOI 10.1074/jbc.RA119.010252

Madoka Kawaguchi1,5, Nurmaa Dashzeveg1, Yue Cao1, Yuzhi Jia1, Xia Liu1, Yang Shen2, and Huiping Liu3

From the 1Department of Pharmacology, Feinberg School of Medicine, Northwestern University, Chicago, Illinois 60611,
2Laboratory of Functional Biology, Graduate School of Biostudies, Kyoto University, Kyoto, 6068501, Japan,
3Department of Electrical and Computer Engineering, TEES-AgrilLife Center for Bioinformatics and Genomic Systems Engineering, Texas A&M University, College Station, Texas 77843, Department of Toxicology and Cancer Biology, College of Medicine, University of Kentucky, Lexington, Kentucky 40536, **Department of Medicine, Hematology/Oncology Division, Feinberg School of Medicine, Northwestern University, Chicago, Illinois 60611, and 3Robert H. Lurie Comprehensive Cancer Center, Feinberg School of Medicine, Northwestern University, Chicago, Illinois 60611

Edited by Eric R. Fearon

CD44 molecule (CD44) is a well-known surface glycoprotein on tumor-initiating cells or cancer stem cells. However, its utility as a therapeutic target for managing metastases remains to be fully evaluated. We previously demonstrated that CD44 mediates homophilic interactions for circulating tumor cell (CTC) cluster formation, which enhances cancer stemness and metastatic potential in association with an unfavorable prognosis. Furthermore, CD44 self-interactions activate the P21-activated kinase 2 (PAK2) signaling pathway. Here, we further examined the biochemical properties of CD44 in homotypic tumor cell aggregation. The standard CD44 form (CD44s) mainly assembled as intercellular homodimers (trans-dimers) in tumor clusters rather than intracellular dimers (cis-dimers) present in single cells. Machine learning–based computational modeling combined with experimental mutagenesis tests revealed that the extracellular Domains I and II of CD44 are essential for its trans-dimerization and predicted high-score residues to be required for dimerization. Substitutions of 10 of these residues in Domain I (Ser-45, Glu-48, Phe-74, Cys-77, Arg-78, Tyr-79, Ile-88, Arg-90, Asn-94, and Cys-97) or 5 residues in Domain II (Ile-106, Tyr-155, Val-156, Gln-157, and Lys-158) abolished CD44 dimerization and reduced tumor cell aggregation in vitro. Importantly, the substitutions in Domain II dramatically inhibited lung colonization in mice. The CD44 dimer-disrupting substitutions decreased downstream PAK2 activation without affecting the interaction between CD44 and PAK2, suggesting that PAK2 activation in tumor cell clusters is CD44 trans-dimer–dependent. These results shed critical light on the biochemical mechanisms of CD44-mediated tumor cell cluster formation and may help inform the development of therapeutic strategies to prevent tumor cluster formation and block cluster-mediated metastases.

CD44 is a multifunctional transmembrane glycoprotein involved in cell–extracellular matrix interaction and cellular signaling cascades (1, 2). The human CD44 gene is located on chromosome 11p13 and consists of 19 exons that encode the full-length CD44 (CD44fl) (2, 4). The first 5 exons contain hyaluronic acid (HA) (4, 5), collagen (7), laminin (2), fibronectin-binding site (8), coding for the extracellular domain. The first 5 and the last 5 exons are constantly transcribed for translation into the standard form (CD44s), whereas the 9 exons located between these regions are subject to alternative splicing, resulting in the generation of CD44 variants (v1 to v10 variants) (9–11). Differential splicing of the 10 variable region exons, as well as variable levels in N-glycosylation and O-glycosylation, result in multiple isoforms of CD44 with different molecular mass (12–15). These wide varieties of variant isoforms and posttranslational modifications are involved in cell–cell and cell–matrix interaction, organ development, neuronal axon guidance, numerous immune functions, lymph node homing, and hematopoiesis (9).

Many cancer types and their metastases express high levels of CD44 (16–18), and CD44 is a well-known surface marker of tumor-initiating cells in breast and other tumors (19, 20). CD44s which is the smallest CD44 isoform, promotes stemness state (21, 22) and is highly expressed in aggressive triple-negative breast cancer cells (23). CD44−CD24−/low cells are thought to be responsible for tumor initiation, metastasis, and therapy resistance (24–26).

We previously reported that, both CD44s and CD44 full-length (CD44fl) are capable of mediating breast tumor cell

1To whom correspondence may be addressed. E-mail: xili369@uky.edu.
2To whom correspondence may be addressed. E-mail: YShen@tamu.edu.
3To whom correspondence may be addressed. E-mail: huiping.liu@northwestern.edu.

This was supported by Susan G. Komen Foundation Grant CCR18548501 (to X. L.) and CCR15332826 (to H. L.); American Cancer Society Grant ACS127951-RSG-15-025-01-CSM (to H. L.); National Institutes of Health Grants TR01CA245699 and R00CA160638 (to H. L.) and R35GM124952 (to Y. S.); DHA, Department of Defense Grant W81XWH-16-1-0021 (to H. L.); and Northwestern University Start-Up Grant (to H. L.). The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the United States Government.

© 2020 Kawaguchi et al. Published under exclusive license by The American Society for Biochemistry and Molecular Biology, Inc.
aggregation, CTC cluster formation, and polyclonal metastases, because of their shared extracellular regions (23). We found that CD44 directly drives cell aggregation through its intercellular, trans-homophilic interactions, which then recruits and activates its downstream target P21-activated kinase 2 (PAK2) (23). Notably, this trans-homophilic CD44 interaction is HA independent in directing tumor cell aggregation. Because CTC clusters possess 23- to 50-fold increased metastatic potential and predict worse prognosis compared with single CTCs (27–29), the biochemical features and the importance of CD44 homophilic interactions in PAK2 activation need to be further investigated.

In this study, to further elucidate the biochemical mechanisms underlying CD44 homophilic interactions, we utilized computational analyses and machine learning–assisted modeling to guide the mutagenesis studies. Based on the relative short sequence of CD44s (encoded by the transcript X4 isoform), we modeled its structure, identified the three extracellular Domains I–III, and further predicted the “hot spot” amino acids in CD44 Domain I and II that are essential for CD44 homophilic dimerization in trans. Mutations of these sites abolished CD44 homophilic interactions and reduced cell aggregation in vitro and/or lung colonization in vivo. Importantly, these mutations also interfered CD44 downstream signaling and decreased PAK2 activation.

**Results**

**Standard CD44 (CD44s) and full-length CD44 (CD44fl) forms both trans- and cis-dimers**

Protein homophilic interactions can take place intercellularly (trans) between neighboring cells or intracellularly (cis) within the same cell. We recently identified that CD44 (CD44s and CD44fl) can form trans-homophilic interactions to mediate CTC cluster formation and drive polyclonal metastases (23). To further determine whether CD44 directs both trans- and cis-homophilic interactions, we designed and conducted two experiments using CD44s-positive MDA-MB-231 breast cancer cells and CD44-negative HEK-293 cells.

First, we observed that upon SDS-PAGE analysis, the endogenous CD44s of 361 amino acids (aa) from the MDA-MB-231 single cell lysate showed at least three monomer forms, of small size (50 kDa (*)), medium size (60 kDa (#)), and big size (80 kDa ($) (Fig. 1A), possibly because of different levels of glycosylation (12–15). To confirm glycosylation of CD44s, we made the CD44s mutant 6Ns/Q (with six Asn residues converted to Gln) with lost N-glycosylation sites. When overexpressed in HEK-

---

**Figure 1. CD44s forms both trans- and cis-dimers.** A, CD44s mainly forms trans-dimers in MDA-MB-231 breast cancer cells. Immunoblots of CD44 and β-actin from MDA-MB-231 single cells and cell clusters, treated with DMSO vehicle control or protein crosslinking agent DSS are shown. The small size, medium size, and big size CD44s monomers are marked as blue (*), pink (#), and black (λ), respectively. Their dimers are labeled as blue (**), pink (##), and black (λλ), accordingly. B, left panel, immunoblots of anti-FLAG (CD44s) of HEK-293 cell lysates transfected with vector control (V), WT CD44s (WT), and mutant CD44 (M = 6Ns/Q) with six Asn residues converted to Gln for depleted glycosylation. Right panel, immunoblots of CD44s showing its lower molecular mass after PNGase F treatment (PN) which partially cleaves the N-glycans from the glycosylated CD44s. NT, no treatment. C, CD44fl forms both trans- and cis-dimers in ectopic CD44fl-overexpressed single HEK-293 cells and clustered HEK-293 cells. @ indicates the main glycosylated monomer CD44fl band. @@ indicate cis- and trans-dimers. D, CD44 intercellular trans-homophilic interactions in clustered HEK-293 cells detected by co-IP. Left panel, diagram of a two-cell cluster, separately transfected with C-terminal FLAG-tagged and HA-tagged CD44, respectively. Right panel, immunoblots for the CD44-FLAG and CD44-HA proteins upon co-IP with anti-FLAG antibodies using lysates from clustered HEK-293 cells after 48-h transfection. The small size, medium size, and big size CD44s monomers are marked as blue (*), pink (#), and black (λ), respectively. E, CD44 intracellular cis-homophilic interactions in single cells detected by co-IP. Left panel, diagram of single HEK-293 cells co-transfected with both FLAG-tagged and HA-tagged CD44. Right panel, immunoblots for the CD44-FLAG and CD44-HA proteins upon co-IP with anti-FLAG antibodies using lysates from single HEK-293 cells after 48-h transfection. The small size, medium size, and big size CD44s monomers are marked as blue (*), pink (#), and black (λ), respectively.
Homophilic CD44 trans-dimerization activate PAK2

293 cells, CD44s mutant 6Ns/Q (CD44s M) showed a molecular mass of ~50 kDa (Fig. 1B, left panel, *). In addition, when using the PNGase F to partially remove N-glycans in MDA-MB-231 cells, CD44s displayed lower molecular mass to ~60 kDa (Fig. 1B, right panel, #). These data suggest that CD44s displays different molecular mass dependent on glycosylated levels.

When the crosslinking reagent disuccinimidyl suberate (DSS) was utilized to stabilize protein complexes, minimal CD44s cis-dimers were detected in the MDA-MB-231 single cell lysates with or without DSS treatment (Fig. 1A). After these single cells aggregated overnight as clusters on the PolyHEMA–coated plate, at least three forms of trans-dimers from three molecular sizes (50 kDa, 60 kDa, and 80 kDa) were increasingly presented at molecular mass of ~100 kDa (**), 120 kDa (##), and 160 kDa (▲▲), respectively, upon DSS treatment (Fig. 1A), suggesting that CD44s favorably form trans-dimers. Furthermore, the CD44s molecules of small size and medium size made the majority of the trans-CD44s dimers whereas the big size CD44s had little dimerization (Fig. 1A).

In the CD44-negative HEK-293 cells transfected with a plasmid encoding the CD44fl of 761 aa, there was one main monomer band at ~110 kDa (Fig. 1C, @), along with the detectable cis-dimers in the single cell lysates as well as trans-dimers (220 kDa) in the clustered cell lysates after DSS cross-linking (Fig. 1C, @@). These data suggest that CD44s and CD44fl might have distinct preferences for cis- and trans-dimers in different cells.

Second, using co-immunoprecipitation (co-IP) we detected CD44 homophilic interactions within HEK-293 cells transfected with tagged CD44s-FLAG and CD44fl-HA. For the trans-dimer analyses, the FLAG- and HA-tagged CD44 plasmids were separately transfected into two different sets of cells, and then these two sets of cells in suspension were mixed and aggregated into clusters prior to cell lysis (Fig. 1D). For cis-dimer analyses, the FLAG- and HA-tagged CD44 plasmids were co-transfected into the same set of cells and then single cells in suspension were harvested for co-IP (Fig. 1E). Consistent with Fig. 1D, the interactions between CD44-FLAG and CD44-HA were detected via co-IP with the anti-FLAG antibody from both experiment designs (Fig. 1, D and E), suggesting that CD44 has an ability to form both intracellular and intercellular homophilic interactions. As our research is mainly interested in CD44-mediated cell aggregation, the following experiments set out to elucidate the biochemical features underlying CD44 intercellular trans-dimerization.

Machine learning–based prediction of the CD44-binding sites for homophilic interactions

CD44 is encoded by 19 exons in human cells and 20 exons in mouse cells with the extra exon 5a (Fig. 2A). As indicated, the first 5 exons of human CD44 encode the N-terminal 223 aa within the extracellular region, which is shared across all the isoforms, and the inserted exons 6–14 generate multiple CD44 variants through alternative splicing (Fig. 2A). CD44s is the smallest CD44 isoform, lacking the entire variable region made by the exons 6–14 (Fig. 2A). Based on the fact that CD44s preferentially mediates trans-dimers in MDA-MB-231 breast cancer cell clusters and the relative better reliability using the shorter sequence for modeling, we focused on modeling CD44s monomers using the webserver iTasser (30) and CD44s dimers using Bayesian active learning (BAL), a machining learning–assisted protein docking method (32).

We identified a four-domain structure of the CD44s monomer with three extracellular Domains I–III and the Domain IV containing the transmembrane and cytoplasmic regions (Fig. 2A). Among the top homodimer models of CD44s, most representative dimer structures indicate the homophilic interactions between the Domains I and II from both monomers. The first-ranked homodimer model suggests head-to-head dimerization from opposite directions at a near 180° angle (Fig. 2B). Interestingly, Domain II of monomer 1 was suggested to dock between Domains I and II of monomer 2. From multistage protein docking, collective analyses of all top dimer models predicted that the hot spot residues visualized in warmer colors are essential for CD44 dimerization, including 12 amino acids (Ser-45, Glu-48, Phe-74, Cys-77, Arg-78, Tyr-79, Ile-88, Arg-90, His-92, Pro-93, Asn-94, and Cys-97) within the Domain I and 5 amino acids (Ile-106, Tyr-155, Val-156, Gln-157, and Lys-158) within the Domain II (Fig. 2C).

Mutant CD44 disrupts trans-dimerization and homophilic interactions

To further determine the importance of these hot spot residues in CD44 homophilic interactions and tumor cell aggregation, we mutated most of these amino acids (except for His-92 and Pro-93) into alanine for functional studies. His-92 and Pro-93 were kept intact to avoid disruption of CD44 protein stability. As listed in Fig. 2C, the rest of the 10 hot spot residues (marked in red) of Domain I were mutated to generate the CD44 DI mutant, and the 5 residues (marked in blue) of Domain II were mutated to make the CD44 DII mutant. The CD44 mutant ΔN21–97 with truncated Domain I (residues 21–97) served as a positive control to disrupt the CD44 trans-homophilic interactions (23). All the three mutants were FLAG-tagged at the C-terminal (same as the WT CD44s-FLAG).

We then investigated whether these predicted sites have impact on CD44 trans-dimerization by transfecting the mutated and WT CD44s (FLAG-tagged) into HEK-293 cells. When the cells were aggregated into clusters, half of the cells were treated with the cross-linker DSS for dimer detection. Although the WT CD44s-FLAG dimerized mainly by the medium size CD44s, the three mutants of CD44s (ΔN, DI, and DII) dramatically disrupted its self-dimerization (Fig. 3A). Furthermore, we utilized a second method of co-IP to evaluate the trans-homophilic interactions between CD44-FLAG and CD44-HA. CD44-FLAG (WT or mutant) and CD44fl-HA were separately transfected into two different sets of HEK-293 cells, and then these two sets of cells were mixed for cell aggregation on Poly-HEMA–coated plate. After 3h clustering, cells were harvested and lysed for co-IP with the anti-HA and the anti-FLAG antibodies. The trans-homophilic interactions between the WT CD44-FLAG and CD44fl-HA proteins were reciprocally detected, respectively (Fig. 3B). However, the FLAG-tagged ΔN21–97, DI, and DII mutants were mostly undetectable in the anti-CD44-HA pulldown complex, and vice versa, CD44-HA was not detected from the anti-FLAG pulldown with
the CD44 mutants (Fig. 3B). These results demonstrate that the CD44 trans-dimerization and homophilic interactions are vitally dependent on the predicted amino acids in Domains I and II. To further identify potential interaction pairs between these amino acids, we conducted machine learning–based CD44 structural modeling and predicted across-chain residue interactions for each hot spot amino acid (Table 1). Many of those predicted interactions are based on opposite charges, aromatic, polar-polar, charge-polar, and hydrophobic-hydrophobic partnerships.

**CD44 mutants block PAK2 activation**

To investigate whether these CD44s mutants have impact on the CD44 downstream signaling, such as the activation of PAK2, we co-transfected PAK2-HA and FLAG-tagged CD44s WT, ΔN21–97, DI, or DII mutants into HEK-293 cells. After 48 h, cells were transferred to Poly-HEMA–coated plates to form clusters. Compared with the CD44s WT, all three CD44 mutants, ΔN21–97, DI, and DII blocked the Ser-20 phosphorylation (activation) of PAK2 (Fig. 4A). Because PAK2 is a cytoplasmic kinase and its interaction with CD44 is predicted to depend on the CD44 cytoplasmic tail, the CD44 extracellular mutants remained in interaction with PAK2 determined by co-IP (Fig. 4B). We further confirmed that knockdown of CD44s in MDA-MB-231 cells decreased PAK2 activation (Fig. 4C). These data suggest that CD44 trans-homophilic interactions are required for PAK2 activation.

**CD44 mutants interfere with cell aggregation in vitro and lung colonization in vivo**

Because the CD44 extracellular mutants abolished CD44 trans-homophilic interactions and its downstream PAK2 activation, we next evaluated their impact on cell-cell aggregation in vitro and lung colonization in experimental metastasis. During the clustering assays on Poly-HEMA–coated plates, cells transfected with ΔN, DI, and DII dramatically decreased the number of larger cell clusters (6–10 cells and >10 cells), compared with the CD44s-transfected HEK-293 cells (Fig. 5, A and B).
Homophilic CD44 trans-dimerization activate PAK2

To determine whether these mutants have an impact on the CTC cluster formation during lung colonization in vivo, we infused the vehicle, WT CD44s or mutant CD44s-transfected MDA-MB-231 CD44 KO cells into the immunodeficient NSG mice via tail vein. We imaged the lung seeding efficiencies of these cells at 24 h post tail vein infusion. Transfection-based WT CD44s expression rescued the lung colonization of CD44s KO cells which was not rescued by the DII mutant (Fig. 6A). Consistently, 48 h after infusion, the KO cells transfected with DII mutant showed significantly fewer big clusters with more than four breast tumor cells (>4 cells) compared with WT CD44s (Fig. 6B). These data suggest that Domain II mutant interferes with lung colonization of tumor cells.

Taken together, our studies have demonstrated that endogenous CD44s prefers trans-homophilic dimerization in breast tumor cells. In addition, combined structural modeling and site-directed mutagenesis revealed that CD44 trans-dimerization and subsequent cell aggregation are dependent on certain residues located within Domains I and II, which also impact PAK2 activation. Furthermore, Domain II mutant significantly decreases lung colonization.

Discussion

Homology modeling, protein docking, and machine learning have shown the great power in the structural modeling of CD44 dimerization in our studies, and they will continue to expedite the CD44-PAK2 structural studies and subsequent drug development. Indeed, machine learning and deep learning have been
applied to exponentially transforming the future technology and increasing the knowledge base, not only in biological research, but also in medicine and other fields.

CD44 is one of the most studied transmembrane glycoproteins in cancer and plays a key role in cell adhesion, signal transduction, and cytoskeleton remodeling (9). It is also a well-known breast tumor-initiating cell marker, and associated with chemoresistance and metastasis (24–26). Most significantly, CD44 has been recently unveiled to directly couple stem cell properties and metastasis through its homophilic interaction.
driven CTC cluster formation and polyclonal metastasis (23). CD44 homophilic interactions in tumor clusters also activate PAK2 signaling pathway that contributes to enhanced stemness and metastatic potential of CTC clusters (23). This finding suggests that targeting CD44 homophilic interactions could potentially interrupt CTC clustering and consequently reduce metastasis. The new identification of the amino acids in Domains I and II essential for CD44 trans-dimerization will facilitate the development of therapeutics that disrupt CTC clustering and therefore block metastasis. Our future studies will further determine the specific contribution of each amino acid in both domains to CD44 homophilic interactions.

Considering that Domains I and II are shared among CD44 isoforms, targeting CD44 homophilic interaction can be a common strategy for all tumor cells with different CD44 variants. Nevertheless, it is also necessary to distinguish tumor cell clustering from CD44-mediated normal cell functions, such as those in lymphocytes. It is worth noting that CD44 shows different bands of molecular mass and its medium-size form dominates the trans-dimerization in tumor cell clusters. This might provide a unique opportunity to identify targetable posttranslational modification features, such as glycosylation patterns of CD44 in tumor clusters for specific drug development and therapeutics.

PAK2 harbors three sites of phosphorylation, and the Ser-20 is known to be phosphorylated by Akt or protein kinase B (36). Phosphorylation of PAK2 at Ser-20 protects cells from apoptosis and promotes cell survival and cellular migration (3, 37). Because the mutations of amino acids in CD44 extracellular domains or deletion of CD44 Domain I did not interrupt CD44–PAK2 interaction, it warrants further identification of the sites mediating CD44 and PAK2 interaction, which most likely locates in transmembrane or cytoplasmic Domain IV. Identification of CD44 regions and amino acids for PAK2 binding could also facilitate the development of peptides to target these sites, interfere with activation of PAK2 and thereby inhibit metastasis.

CTCs are cells that have detached from primary tumor and entered the bloodstream. Although extraordinarily rare, they are considered the seeds for the subsequent growth of metastasis in distant organs, especially CTC clusters. In addition to CD44-mediated CTC cluster aggregation, there might exist CD44-independent mechanisms underlying CTC cluster formation to be targeted for drug development (23–25). By screening FDA-approved compounds, Na+ / K+ ATPase inhibitors were identified as potential drugs to suppress metastasis by targeting CTC clusters (29). In vivo administration data of these inhibitors suggest that CTC cluster prevention should be implemented at the time of localized disease and before tumor dissemination into distant sites. Meanwhile, if similar mechanisms support the survival and aggregation of CTC clusters before extravasation, targeting CTC clusters could also result in therapeutic benefits for patients with late-stage disease. However, given the time from cancer cells entering into circulation to reaching a secondary organ is a relatively short time, it is equally important to directly target clustered CTCs and micrometastases in distant organs in addition to preventing tumor cell aggregation prior to extravasation as well as CTC cluster formation in circulation. A better understanding of the upstream and downstream signaling pathways underlying CTC cluster formation could help control and block cancer metastasis.

Materials and methods

Animal studies

All animal procedures and experimental procedures have been performed under approval by Northwestern University
Animal Care and Use Committee (ACUC) and complied with the National Institutes of Health Guidelines for the Care and Use of Laboratory Animals. All mice used in this study were kept in specific pathogen-free facilities in the Animal Resources Center at Northwestern University. 8- to 10-week-old female NSG mice were used for tail vein injection of human cells. The patient-derived xenograft tumors were established as described previously (18, 24).

**Tail vein injection and lung dissection**

For tail vein injection, L2G-labeled 231-CD44KO cells over-expressing CD44s-FLAG, CD44s-Domain I mutant-FLAG (DI), and CD44s-Domain II mutant-FLAG (DII) in 200 μl of PBS were injected into NSG mice via tail vein using the 28-gauge syringe needles (BD Biosciences, 329461). 48 h post tail vein injection, the mice were euthanized for tissue dissection. After washing with cold PBS, the dissected lungs were imaged under fluorescence microscopy.

**Bioluminescence imaging**

After intraperitoneal injections with 100 μl of D-luciferin (30 mg/ml, Gold Biotechnology, 11 5144-35-9), mice were anesthetized with isoflurane. Bioluminescence images were acquired on day 0, 1, and 2 using the Spectral Instrument Imaging Spectral Lagos (SII LAGO) and the signals were presented as total flux (photons/second) and -fold changes of the total flux signals (Aura, version 2.2.1.1). Acquisition times ranged from 5 s to 5 min.

**Cell lines**

Human breast cancer cell line MDA-MB-231 and an immortalized HEK-293 cell were purchased commercially from American Type Culture Collection. L2G-labeled 231-CD44KO cells were established as described previously (23). Early passage of cells (<20 passages) were maintained in DMEM (Fisher Scientific, SH30243FS) supplemented with 10% FBS (Takara Bio, 631106) and 1% penicillin-streptomycin (Sigma-Aldrich, 28-gauge syringe needles (BD Biosciences, 329461). 48 h post tail vein injection, the mice were euthanized for tissue dissection. After washing with cold PBS, the dissected lungs were imaged under fluorescence microscopy.

**Transfections**

CD44s-ΔNd21_97-FLAG, CD 44s-Domain I mutant-FLAG (DI), CD44s-Domain II mutant-FLAG (DII), and CD44s-6Ns/Q mutant for de-N-glycosylation (six Asns, including Asn-25, Asn-57, Asn-100, Asn-101, Asn-110, and Asn-120, converted to Gln) were made by mutagenesis (GenScript) based on pCMV6-FLAG-CD44. For plasmid-cDNA overexpression experiments, pCMV3-C-GFP Spark Control Vector (Sino Biological, CV026), pCMV3-HA-PAK2 (Sino Biological, HG10085-NY), pCMV3-CD44-HA (Sino Biological, HG12211-CY), pCMV6-FLAG-CD44 (OriGene, RC221820), CD44s-ΔNd21–97-FLAG, CD44s-Domain I mutant-FLAG (DI), and CD44s-Domain II mutant-FLAG (DII) were transfected into HEK-293 cells using PolyJet (SigmaGen Laboratories, SL100688). At 48 h post transfection, cells were trypsinized and filtered through 35-μm nylon mesh to make single cell suspension and lysates. For cluster formation, single cells aggregate upon incubation on the Poly-HEMA (Sigma-Aldrich, P3932-10G)–coated dish for 3–6 h. Collected single and clustered cells were used for crosslinking assay, co-IP, and Western blotting. L2G-labeled MDA-MB231 cells–CD44KO cells were double transfected with plasmid-cDNA overexpression using JetPRIME (VWR, 89129-920). For siRNA transfection, 100 nM siRNAs against CD44 (Dharmacon, L-009999-00-0010), PAK2 (Dharmacon, L-003597-00-0020), and negative control (Dharmacon, D-001810-10^-50) were transfected using Dharmafect (Dharmacon, T-2001-03) according to the manufacturer’s protocol. On the following day, medium was changed and cells were retransfected for the second time prior to be collected for analyses.

**PNGase F treatment**

For removing N-linked oligosaccharides from CD44s, used PNGase F according to manufacturer’s protocol (New England Biolabs, P0704S). Briefly, MDA-MB-231 cell lysates were first denatured with provided denature buffer for 10 min and incubated with 1000U of PNGase F to cleave N-glycans.

**co-IP and Western blotting (WB)**

After washing twice in cold PBS, cells were lysed in RIPA buffer (Cell Signaling Technology, 9806S) or IP lysis buffer (Thermo Fisher Scientific, 87787) with protease inhibitor mixture (Thermo Fisher Scientific, 78440) and incubated for 30 min on ice. Protein concentration was determined by Protein Assay Dye Reagent (Bio-Rad, 500-0006). For co-IP, equal amounts of protein of each sample were incubated with anti-HA Magnetic Beads (Thermo Fisher Scientific, 88836) or anti-FLAG M2 Magnetic Beads (Sigma-Aldrich, M8823-5ML) overnight at 4 °C. For elution, 0.1 mg/luciferin (pH 2–3) was added and incubated for 5–10 min at room temperature prior to the supernatant collection. Then the same amount of Triton X-100 (TBS) was added for neutralizing the low pH. For WB, equal amounts of protein were diluted in 10% β-mercaptoethanol-contained SDS-PAGE loading buffer (Bio-Rad 4Laemmli Sample Buffer 161-0747). Proteins were resolved by size using 4–20% or 7.5% SDS-PAGE gel, and then transferred to 0.2-μm pore size nitrocellulose membranes. Blocking was with 5% BSA in TBS/0.1% Tween 20 for 1 h at room temperature and incubated with primary antibodies for 1 h at room temperature or 4 °C overnight. The following primary antibodies and dilutions were used: CD44 (1:1000; Thermo Fisher Scientific, MA513890), FLAG (1:1000; Sigma-Aldrich, F1804-200UG), HA (1:500; Santa Cruz Biotechnology, 12CA5), PAK2 (1:1000, Thermost Fisher, PA5-20254), phospho-PAK2 (1:1000, Cell Signaling Technology, 2607T), β-actin (1:5000, Abcam, ab8224). After washing primary antibodies in TBS/0.1% Tween 20, membrane was incubated with horseradish peroxidase (HRP)-conjugated secondary antibodies (mouse: 1:10,000, Promega, 402B; rabbit: 1:10,000, Promega, W401B) for 1 h at room temperature. HRP-conjugated secondary antibodies were detected by enhanced chemiluminescence (Thermo Fisher Scientific, 32132).

**Crosslinking assay**

After two washings in cold PBS, collected cells were treated with 2.5 mM of crosslinking reagent DSS (Thermo Fisher, 21655) for 2 h on ice. Then the quench solution (1 mM Tris-HCl...
Homophilic CD44 trans-dimerization activate PAK2

pH 7.5, 1:100 dilution) was added to a final concentration of 10 mM for 15 min on ice. Finally, the cells were lysed with RIPA buffer for WB or IP lysis buffer for co-IP.

Cell clustering assay

After 48 h post transfection, cells were trypsinized and filtered through 35-μm nylon mesh to make single cell suspension. Cells were suspended and incubated onto Poly-HEMA-coated 24-well plates. Cells were observed under microscope for 4 h at room temperature with images taken at various time points.

Structural modeling

The 3D structure of CD44 antigen derived from the transcript isoform 4 precursor (NM_001001391) was first built using the webserver iTasser (30). It was then rigidly docked to itself using the webserver ClusPro (31) under the homodimer mode. The top 10 initial models under the default ClusPro scoring function were refined using BAL (32). BAL describes protein flexibility and motions in complex normal modes from cNMA (33, 34), adopts a scoring function derived from machine learning, samples conformations using a novel Bayesian active learning method, and estimates the conditional probability, the uncertainty, and the quality range of each model. The refined models were re-ranked based on their estimated conditional probabilities of being near-native (interface root mean squared deviation not exceeding 4Å). Furthermore, each residue was assigned a probability of being at the dimer interface by weighting models by their conditional probabilities and averaging the binary indicators for putative interfaces across all models. The maximum value of the residue probability is 1, indicating that the corresponding residue is surely at the interface given that all structural models contain at least one near-native dimer structure. Similarly, each across-chain residue pair was assigned a probability of being in contact (if any pair of heavy atoms are within 5Å). The protein structures and dimerization hot spots were visualized using the computer program PyMol (35).

Statistical analysis

For image analysis, images were taken at least five random fields of the clustering cells and the fluorescence lungs images from three mice. A two-tailed Student’s t test was used to evaluate the p values, and p < 0.05 was considered statistically significant and represented with one asterisk (*). Probabilities under 0.01 were represented with two asterisk (**), and under 0.001 represented with three asterisk (***) . Data are presented as mean ± S.D.

Acknowledgments—Portions of this research were conducted with high performance research computing resources provided by Texas A&M University.

References