

Molecular basis of sidekick-mediated cell-cell adhesion and specificity

Kerry M. Goodman^{1,8}, Masahito Yamagata^{6,8}, Xiangshu Jin^{1,3,7}, Seetha Mannepalli¹,

Phinikoula S. Katsamba^{1,2,3}, Göran Ahlsén^{2,3}, Alina P. Sergeeva^{2,3},

Barry Honig^{1,2,3,4,5,*}, Joshua R. Sanes^{6,*}, Lawrence Shapiro^{1,2,5,*}

¹Department of Biochemistry and Molecular Biophysics, ²Department of Systems Biology, ³Howard Hughes Medical Institute, ⁴Department of Medicine, ⁴Zuckerman Mind Brain and Behavior Institute, Columbia University, New York, NY 10032, USA.

⁶Department of Molecular and Cellular Biology and Center for Brain Science, Harvard University, Cambridge, MA 02138, USA.

⁷Present address: Department of Chemistry, Michigan State University, East Lansing, MI 48824, USA.

⁸Co-first author

*Correspondence: bh6@cumc.columbia.edu (B.H); sanesj@mcb.harvard.edu (J.R.S.); shapiro@convex.hhmi.columbia.edu (L.S.)

Summary

Sidekick (Sdk) 1 and 2 are related immunoglobulin superfamily cell adhesion proteins required for appropriate synaptic connections between specific subtypes of retinal neurons. Sdks mediate cell-cell adhesion with homophilic specificity that underlies their neuronal targeting function. Here we report crystal structures of Sdk1 and Sdk2 ectodomain regions, revealing similar homodimers mediated by the four N-terminal immunoglobulin domains (Ig1–4), arranged in a horseshoe conformation. These Ig1–4 horseshoes interact in a novel back-to-back orientation in both homodimers through Ig1:Ig2, Ig1:Ig1 and Ig3:Ig4 interactions. Structure-guided mutagenesis results show that this canonical dimer is required for both Sdk-mediated cell aggregation (via *trans* interactions) and Sdk clustering in isolated cells (via *cis* interactions). Sdk1/Sdk2 recognition specificity is encoded across Ig1–4, with Ig1–2 conferring the majority of binding affinity and differential specificity. Competition between *cis* and *trans* interactions, which form through the same molecular interface, could sharpen the specificity of cell-cell interactions.

Introduction

In the vertebrate retina, light-sensitive photoreceptors synapse on interneurons; these interneurons process the information and pass it to retinal ganglion cells (RGCs), which send it to the brain (Masland, 2012). Highly stereotyped patterns of connectivity between the ~70 types of interneurons and ~30 types of RGCs render the latter sensitive to specific visual features such as motion or edges (Sanes and Masland, 2015). Synapses between these interneurons and RGCs form in the inner plexiform layer (IPL) of the retina, with arbors of each specific neuronal subtype confined to one, or a few, of the approximately 10 sublaminae (Roska and Werblin, 2001; Sanes and Zipursky, 2010).

Some aspects of this specific connectivity appear to be mediated by recognition molecules of the immunoglobulin superfamily (IgSF). Studies in chicks and mice have revealed that defined interneuron and RGC subtypes express one or more of 10 closely related IgSF members: Sdk1, Sdk2, Dscam, DscamL1, and Contactins 1–6 (CNTN1–6) in largely non-overlapping patterns (Yamagata et al., 2002; Yamagata and Sanes, 2008, 2012a; Fuerst et al., 2008, 2009; Shekhar et al., submitted). In chick, Sdk, Dscam and CNTN family proteins are present as interneuron-RGC synapses form, and both knockdown and over-expression experiments show that they are necessary and sufficient for directing neural processes to particular sublaminae in the IPL (Yamagata et al., 2002; Yamagata and Sanes, 2008, 2012a). In mice, Sdk1, Sdk2, Dscam, DscamL1 and CNTN5 mutants each exhibit specific defects in arborization and connectivity within the IPL (Fuerst et al., 2008, 2009; Krishnaswamy et al., 2015; Y. Peng, N. Tran and J.R.S., unpublished). In one case, the specific connectivity of an interneuron type (vesicular glutamate transporter 3-positive amacrine cells or VG3-ACs) to a specific RGC type (W3B-RGCs) depends upon expression of Sdk2 in both cell types: transmission from VG3-ACs to W3B-RGCs fails in Sdk2 mutants and the RGCs no longer respond to their canonical visual feature (Krishnaswamy et al., 2015). These results have led to the hypothesis that IgSF-mediated homophilic interactions bias synaptic connectivity in favor of appropriate partners, thus generating information processing circuits in the retina. Since all 10 of these IgSF molecules are also expressed by neuronal subsets throughout the central nervous system

(Yamakawa et al., 1998; Agarwala et al., 2001; Shimoda and Watanabe, 2009; Stoeckli, 2010; Yamagata and Sanes, 2012a), similar interactions may mediate connectivity in multiple brain regions.

Sdk1 has also been shown to be involved in the pathology of focal segmental glomerulosclerosis and HIV-associated neuropathy (Kaufman et al., 2004, 2007, 2010). Inappropriate up-regulation of Sdk1 expression by podocytes has been linked to their dedifferentiation and loss of proper foot-process architecture, leading to collapsed glomeruli and neuropathy (Kaufman et al., 2007). Sdk1 is expressed at high levels during kidney development, with very low expression afterwards. Sdk1-associated kidney pathologies are thought to reflect a reversion of podocytes to the early developmental state, caused by inappropriate Sdk1 expression (Kaufman et al., 2004, 2007, 2010).

Sdk1 and Sdk2 are single-pass transmembrane proteins, with extracellular regions composed of 6 N-terminal immunoglobulin (Ig) domains followed by 13 fibronectin type III (FNIII) domains, and a relatively short intracellular domain terminating in a Postsynaptic density/Discs Large/ZO-1 (PDZ) binding motif (Figure 1A) (Nguyen et al., 1997; Yamagata et al., 2002; Yamagata and Sanes, 2010; Kaufman et al., 2010). Binding of this C-terminal motif to scaffolding molecules of the membrane-associated guanylate kinase with inverted orientation (MAGI) family is necessary for synaptic localization of Sdks, and required for appropriate function in the retina and kidney (Yamagata and Sanes, 2010; Kaufman et al., 2010).

IgSF neural recognition proteins from the Dscam, CNTN and L1 families have extracellular domain architecture related to the Sdks: Dscam and DscamL1 have 10 extracellular Ig and 6 FNIII domains (Schmucker and Chen, 2009); CNTNs contain 6 Ig and 4 FNIII domains (Shimoda and Watanabe, 2009); and L1-related molecules have 6 Ig and 4 or 5 FNIII domains (Maness and Schachner, 2007). The four N-terminal Ig domains are arranged in a horseshoe conformation in Dscams, CNTNs, and L1-family proteins (Chen et al., 2013). Despite this similarity in protomer architecture, crystallographic studies of the *Drosophila* Dscam ortholog, Dscam1 (Meijers et al., 2007; Sawaya et al., 2008), human CNTN2

(AXONIN-1/TAG-1) (Mortl et al., 2007), mouse CNTN4 (Bouyain and Watkins, 2010), and the human L1 family member Neurofascin (Liu et al., 2011), revealed distinct homodimer structures mediated by horseshoe motifs.

Here we report crystal structures of cell-cell adhesive homophilic dimers of mouse Sdk1 and Sdk2, each mediated by the four N-terminal Ig domains. These four domains adopt a horseshoe conformation, like many other IgSF cell-cell recognition proteins, but they interact in a unique back-to-back anti-parallel manner not previously observed. Mutagenesis studies both *in vitro*, with analytical ultracentrifugation (AUC) and surface plasmon resonance (SPR) readouts, and *in situ*, with a cell aggregation assay readout, demonstrate that the crystallographic dimer is present in solution and is required for Sdk-mediated cell aggregation. Interestingly, this same dimer is also required for *cis*-clustering of Sdk-molecules in isolated cells. Structures of multiple crystal forms of Sdk1 and Sdk2 revealed an unexpected flexibility in the dimer arrangement: The antiparallel contacts between Ig1–2 regions from each protomer were maintained in all structures, however the degree of contact between the Ig3–4 regions was highly variable. Consistent with this observation, mutagenesis studies showed that the Ig1–2:Ig1–2 interaction was necessary for dimerization, whereas mutations that interfered with the Ig3–4:Ig3–4 interaction had only a modest effect on dimer affinity. Overall, our data suggest a model in which Sdks form *cis* dimers on isolated cell surfaces, which dissociate to form *trans* dimers through the same interface when contact is made to a cell surface expressing the cognate Sdk. Competition between these *cis* and *trans* dimers may provide a mechanism to enhance the homophilic specificity of Sdk-mediated interactions.

Results

The adhesive Sidekick dimer is mediated by Ig1–4

Consistent with their role in defining neuronal contacts, both Sdk1 and Sdk2 mediate homophilic adhesion when applied to beads or transfected into cultured cells (Yamagata et al., 2002; Yamagata and Sanes, 2008; Figure 1). A chimeric construct (SdkD, Figure 1A) comprising Ig1–5 and part of Ig6 from

Sdk2 and the remainder of the molecule from Sdk1 could mediate adhesion to Sdk2 but not Sdk1 in a mixed cell aggregation assay, using either L cells (Figures 1B and 1C) or N-cadherin deficient HEK-293 cells (data not shown), indicating that it is the Ig domain region that mediates cell-cell recognition in common with other IgSF proteins (Gouveia et al., 2008; Haspel et al., 2000; Liu et al., 2011; Wojtowicz et al., 2004; Sawaya et al., 2008). We also asked whether the cytoplasmic domain is required for cell-cell adhesion. To this end, we replaced the cytoplasmic domains of Sdk1 and Sdk2 with fluorescent proteins. Adhesion was unperturbed by this replacement (Figure 1D). Thus Sdk-mediated cell-cell adhesion requires the extracellular but not the intracellular domains of the proteins, with key determinants of homophilic specificity in Ig1–6.

To further define and measure the adhesive interaction for mouse Sdk1 and Sdk2, we produced soluble Ig1–4, Ig1–5 and Ig1–6 constructs in HEK-293 cells. Sedimentation equilibrium analytical ultracentrifugation (AUC) measurements showed that Sdk1 and Sdk2 Ig1–4, Ig1–5, and Ig1–6 were each dimers in solution with low-micromolar affinities (Table 1) with the Sdk2 dimer exhibiting ~5-fold stronger affinity than the Sdk1 dimer for each truncation construct tested. These affinities are similar to other cell-cell recognition proteins, such as *Drosophila* Dscam1 isoforms (1–2 μ M; Wu et al., 2012) and classical cadherins (8–130 μ M; Harrison et al., 2011; Vendome et al., 2014). Ig1–4 is therefore sufficient for dimerization in solution for both Sdks.

The four N-terminal Ig domains are arranged in a stable horseshoe conformation

To determine the nature of the adhesive interaction for both Sdk1 and Sdk2, we determined crystal structures of Sdk1_{Ig1–4}, Sdk1_{Ig1–5}, Sdk2_{Ig1–4}, and an Sdk2_{Ig1–2}/Sdk1_{Ig3–4} chimera (Figure 2A). Two crystal forms of Sdk1_{Ig1–4} were determined at 2.2 and 3.2 Å resolution respectively, one of Sdk1_{Ig1–5} at 3.5 Å, two of Sdk2_{Ig1–4} at 2.7 and 3.2 Å respectively, and one of the Sdk2_{Ig1–2}/Sdk1_{Ig3–4} chimera at 2.7 Å resolution. Data collection and refinement statistics are given in Table S1.

The N-terminal four Ig domains of both Sdk1 and Sdk2, and the chimeric construct, Sdk2_{Ig1-2}/Sdk1_{Ig3-4}, are arranged in highly similar horseshoe structures (Figure 2A), with pairwise root mean square deviations over aligned C α atoms (RMSDs) among all protomers of 2.4 Å or less.

The horseshoe conformation is formed by an anti-parallel interaction between Ig1-2 and Ig3-4, which are connected by an eight amino acid linker between Ig2 and Ig3 (Sdk1 R185-A192 and Sdk2 N185-P192). Two non-overlapping interfaces hold the horseshoe together: a conserved and relatively rigid Ig1:Ig4 interface and a more flexible and varied Ig2:Ig3 interface (Figures 2B-D and Table S3). These intramolecular interactions bury extensive combined intramolecular surface areas of 2620 Å² in Sdk1, 2459 Å² in Sdk2 and 2567 Å² in the Sdk2_{Ig1-2}/Sdk1_{Ig3-4} chimera, implying that the horseshoe should be a stable element in all of these molecules. In this conformation, Ig1 and Ig2 are arranged in tandem with a ~40° bend between them; Ig3 and Ig4 are also arranged in tandem with a ~30° inter-domain bend (Table S2). Horseshoe conformations are also found for the N-terminal four Ig domains of a number of other IgSF molecules including all members of the deleted in colorectal cancer (DCC), CNTN, L1, Dscam, cell adhesion associated oncogene regulated (CDON), MAM domain-containing glycosylphosphatidylinositol anchors (MDGA), and activated leukocyte cell adhesion molecule (ALCAM) families (Chen et al., 2013). Crystal structures of the horseshoe regions are available for *Drosophila* Dscam1 (Meijers et al., 2007; Sawaya et al., 2008) and vertebrate CNTN (Freigang et al., 2000; Mortl et al., 2007; Bouyain and Watkins, 2010), DCC (Chen et al., 2013), and L1 family members (Liu et al., 2010; Su et al., 1998). Comparison between each of these structures and Sdk reveals the horseshoe conformation is similar overall (RMSDs from 2.8 to 6.2 Å; Table S3) despite sequence identities with Sdk1 of only 20–26%.

The internal Ig1:Ig4 interface in both Sdk1 and Sdk2 includes conserved ‘horseshoe motif’ hydrogen bonds, identified by Chen et al. (2013) from their analysis of a number of horseshoe proteins. These occur between the side chains of Ig4 residues D325 (from Ig4 C strand WXXN/D motif) and Q359 (from Ig4 F strand Φ Y/F/LQC motif) and the main chain of Ig1 residues A79 and L81 (Figures 2D and Figure 2–figure supplement 1). Despite the presence of these motifs and the requisite long Ig2–Ig3 linker, Chen et

al., who surveyed the presence of IgSF horseshoe motifs through sequence analysis, did not identify Sdks as horseshoe motif-containing molecules. This is because Sdks' lack an internal disulfide in Ig2, which Chen et al. used to define the Ig2–Ig3 linker length in their bioinformatics search. The internal Ig2:Ig3 interface appears to lack conserved features amongst IgSF horseshoe proteins (Chen et al., 2013). In Sdk1 and Sdk2 the Ig2:Ig3 interaction is mediated by main chain to side chain hydrogen bonds between Sdk-conserved residues (S137_{Ig2}:R247_{Ig3}, I142_{Ig2}:R248_{Ig3}, and L144_{Ig2}:E216_{Ig3}), alongside van-der-Waals interactions among hydrophobic residues (L144_{Ig2}, V149_{Ig2}, L151_{Ig2}, V/I198_{Ig3}, V/I199_{Ig3}, and V218_{Ig3}) (Figure 2C). Remarkably, the Sdk2_{Ig1–2}/Sdk1_{Ig3–4} chimera structure features the same set of hydrogen bonds and hydrophobic interactions that mediate both Ig1:Ig4 and Ig2:Ig3 interfaces.

The Sdk1_{Ig1–5} structure showed that Ig5 extends laterally from the horseshoe with a 54° deviation from linearity with Ig4 (Figure 2A). This angle is maintained in both independent Ig1–5 chains observed in the crystal structure. The linker between Ig4 and Ig5 is only one amino acid (N379) and there is a small Ig4:Ig5 interface with a buried surface area (BSA) over both domains of ~370 Å². This interface—involving the linker region, the Ig4 AB loop and the Ig5 BC loop—likely provides rigidity to the Ig4–Ig5 junction. In Sdk2 the Ig4–Ig5 linker is the same length as in Sdk1, although the linker residue is serine rather than asparagine, and the Ig4:Ig5 interfacial residues are mostly conserved between Sdk1 and Sdk2 (Figure 4—figure supplement 1), except key Ig4 AB loop residue 302 which is a methionine in Sdk2 rather than a valine. The arrangement of Ig4 and Ig5 in Sdk2 is therefore likely to be similar to that observed in Sdk1. However the surface of the Ig4–Ig5 linker region in Sdk1 is highly acidic, whilst that of Sdk2 is comparably neutral (Figure 4—figure supplement 2).

We were unable to obtain crystals of an Sdk1 or Sdk2 construct containing Ig6 that diffracted to sufficient resolution for structure determination. The Ig5–Ig6 linker is ~2–4 amino acids long for both Sdk1 and Sdk2 (Figure 4—figure supplement 1), which could potentially accommodate a range of Ig5–Ig6 bend angles. The positioning of Ig6 is therefore unknown, however since Ig5 projects away from the dimer

interface it is unlikely Ig6 would be able to contribute to the dimer observed in the crystal structures, which is described in detail below.

The Sidekick dimer is mediated by a flexible back-to-back interaction between horseshoes

Consistent with the AUC results of the wild-type Sdk1 and Sdk2, the crystal structures show a dimeric arrangement of molecules, consisting of symmetrical back-to-back interactions (convex-face to convex-face) between the Ig1–4 horseshoe regions of the individual protomers, with Ig5 making no contacts in the Sdk1_{Ig1–5} dimer (Figure 3A). The dimer protomers are related by crystallographic 2-fold symmetry in both of the Sdk1_{Ig1–4} structures, and by non-crystallographic symmetry in all the other structures. The Sdk1 and Sdk2 dimers are predominantly mediated by symmetrical anti-parallel interactions between the Ig1–2 halves of the horseshoes, with a contribution from an anti-parallel interaction between the Ig3–4 halves of the horseshoes observed in some, but not all of the crystal forms (Figures 3B and 3C).

The Ig1–2:Ig1–2 interactions in both the Sdk1 and Sdk2 structures consist of an Ig1:Ig1 interface near the dimer two-fold axis, and two symmetry-related Ig1:Ig2 interfaces. The Ig1:Ig1 interface, for both Sdk1 and Sdk2, is centered on N22, which hydrogen bonds with the main chain of R23, and is supported by interactions between hydrophobic residues (L19, V25) (Figures 4A and 4B). The Sdk1 and Sdk2 Ig1:Ig2 interfaces, which are contiguous with the central Ig1:Ig1 interface, consist of two networks of hydrogen bonding interactions clustered around residues E31_{Ig1} and E/D168_{Ig2}; a salt bridge between E31_{Ig1} and K133_{Ig2} (corresponding to a region of complementary electrostatic potentials between Ig1 and Ig2, Figure 4–figure supplement 2); and a number of hydrophobic residue contacts including V4_{Ig1}:I/P135_{Ig2} and L/M29_{Ig1}:V166_{Ig2} (Figures 4A and 4B). The Ig1:Ig1 and Ig1:Ig2 interfaces of Sdk1 and Sdk2 are remarkably similar, with only five relatively conservative differences in the identity of the residues involved: I/V17_{Ig1} (marginally interfacial), L/M29_{Ig1}, K/R55_{Ig1}, I/P135_{Ig2}, and E/D168_{Ig2} (Figures 4A–C and Figure 4–figure supplement 1). There are also conserved differences in Ig1 residues 13–15 (GLP in Sdk1 and VRT in Sdk2; Figure 4–figure supplement 1), which form a loop near, but not within, the

Ig1:Ig1 interface. The role of interacting variable residues 29_{Ig1} and 168_{Ig2} in Sdk1/Sdk2 specificity is discussed below.

Although the Ig1–2:Ig1–2 interaction is present in all of the crystal forms of Sdk1 and Sdk2, there is a difference in rotational angle between the two horseshoes of the dimer among the crystal forms, with the Ig1–2:Ig1–2 interface acting as the hinge (Figures 3B and 3C). This results in differing levels of interaction between the Ig3–4 regions in the Sdk1 and Sdk2 dimers in the different crystal forms (Figure 1H). The Ig3–4:Ig3–4 interaction varies from the Ig3–4 domains forming a considerable additional interface—as in Sdk1_{Ig1–4} crystal form 2 (915 Å² BSA) and Sdk2_{Ig1–4} crystal form 1 (1257 Å² BSA)—to being splayed 14 Å apart, as in the Sdk1_{Ig1–4} crystal form 1 structure (0 Å² BSA). These differences indicate that there is flexibility in the interaction between the horseshoes in solution, with different conformations being trapped in the crystals. These observations raise questions about how much the Ig3–4:Ig3–4 interfaces contribute to dimer stability and whether in cells this flexibility allows the accommodation of an as yet unidentified molecule in the Ig3–4:Ig3–4 clefts.

Analysis of the Ig3–4:Ig3–4 interface regions in the Sdk1_{Ig1–4} crystal form 2 and Sdk2_{Ig1–4} crystal form 1 structures, which show the most extensive Ig3–4:Ig3–4 contacts, reveals predominantly hydrophilic surfaces with hydrogen bonds between N253_{Ig3} and P287_{Ig4} in both structures, S240_{Ig3} and T291_{Ig4} in Sdk1, and T253_{Ig3} and E286_{Ig4} in Sdk2 (Figures 4A and 4B). The Ig3–4:Ig3–4 interface is not as highly conserved as the Ig1–2:Ig1–2 interface even within Sdk1 and Sdk2 homologs (Figure 4C and Figure 4—figure supplement 1). However, there are three Ig3–4:Ig3–4 interfacial residues that show non-conservative differences between Sdk1 and Sdk2 that are mostly conserved amongst vertebrate orthologs (H/S243_{Ig3}, Y/Q289_{Ig4} and S/R296_{Ig4}) (Figure 4C). These differences result in the Sdk2 interface containing a π -stacking interaction between Y245_{Ig3} and R296_{Ig4}, which is absent from Sdk1 (Figures 4A and 4B). Similarly, the Sdk1 interface contains an H243_{Ig3}:E293_{Ig4} interaction, which is absent from Sdk2, although the contribution of this interaction is likely small since the side chains are still 4.4 Å apart in the Sdk1 crystal structure with the most extensive Ig3–4:Ig3–4 interface (Figure 4A).

In addition to these major interfacial regions, the Sdk1_{Ig1-4} crystal form 2 also shows a small contact region between Ig1 residues E11 and R83 and Ig3 residue E212 (Figure 4A) towards the center of the horseshoe dimer. This interaction is not seen in either of the Sdk2 structures, although the residues are conserved.

Biophysical analysis of Sdk mutants reveal the relative contributions of the Ig1-2 and Ig3-4 regions to dimerization

We performed mutagenesis experiments to assess the contribution of the various contact regions in the crystallographically defined dimer. Mutating residues in the central Ig1:Ig1 interface (N22R in Sdk1 and H18R/N22S in Sdk2) resulted in the loss of dimerization for both Sdk1 and Sdk2 in AUC experiments (Table 1). Consistent with this, a crystal structure of Sdk2_{Ig1-4} H18R/N22S showed a loss of the back-to-back dimer interaction (Figure 4-figure supplement 3). Additionally, mutation of a salt-bridge residue (K133E) from the Ig1:Ig2 interface in Sdk1 resulted in 20-fold weaker binding in solution (Table 1).

Mutagenesis of a key Ig3-4:Ig3-4 interface residue (N253E) did not prevent Sdk dimerization, although it did reduce the dimer affinity, particularly for Sdk2, which went from 2.2 μ M to 18.9 μ M, implying a significant albeit limited Ig3-4 contribution to dimer strength. Interestingly the Sdk2_{Ig1-2}/Sdk1_{Ig3-4} chimera had a dimerization affinity intermediate to that of Sdk1 and Sdk2 (Table 1), suggesting that the Sdk2 Ig1-2 contributions to dimer affinity may be more than Sdk1 Ig1-2 while the Sdk1 Ig3-4 contributions to dimer affinity may be less than Sdk2 Ig3-4, and could therefore both underlie the lower dimer affinity of Sdk1 relative to Sdk2. The crystal structure of this chimera did not show an interaction between Ig3-4 regions in the dimer, although given the variability observed in the conformations of the dimers in the different Sdk1 and Sdk2 crystal structures, the Ig3-4 regions may interact in solution (Figures 3B and 3C).

The Ig1–4 dimer is required for Sdk1-mediated cell adhesion

To determine whether the crystallographic dimer is the adhesive Sdk dimer, we tested the N22R mutation in the context of full length Sdk1 to see whether it would prevent Sdk-mediated cell adhesion in a cell aggregation assay. Aggregation of cells expressing the Sdk1 N22R mutant was significantly reduced compared to those expressing wild-type Sdk1 (Figures 5A–B for HEK cells and Figure 5–figure supplement 1 for L cells). In addition Sdk1 N22K and L29E mutants also significantly reduced cell adhesion and the Sdk1 K133E mutant showed decreased cell aggregation in the same assay (Figure 5B). We also analyzed distribution of Sdk1 at sites of cell-cell contact. Wild-type Sdk1 localized to the cell-cell junctions, whereas the Sdk1 N22R mutant showed diffuse localization over the cell surface (Figure 5C) and showed no increased localization at sites of contact with cells that expressed wild-type Sdk1 (Figure 5D). These data indicate that the dimer observed both in solution and in the crystal structures is necessary for Sdk-mediated cell adhesion. The simplest conclusion from these data is that this dimer represents the adhesive dimer formed between molecules emanating from opposing cells (*trans* interaction).

Sdk1 forms cis clusters dependent on the Ig1-4 dimer

The clustering of Sdk1 at sites of cell-cell contact is consistent with its role in intercellular adhesion. However, we found that Sdk1 also formed clusters on solitary isolated cells in the same cultures (Figure 5E). This clustering was not seen in cells expressing the Sdk1 N22R mutant; in these cells, Sdk1 N22R was diffusely localized (Figure 5E). These observations demonstrate that Sdk1 forms clusters *in cis*, which are not dependent upon *trans* (cell-cell) interaction, but are dependent upon the identified horseshoe-mediated dimer interaction. This suggests that the identified dimer can form *in cis*.

To assess the specificity of the *cis*-interaction, we generated constructs in which Sdk1, Sdk1 N22R or Sdk2 were fused to a fluorescent protein (RFP or YFP). We then clustered the Sdk1-RFP using a co-transfected nanobody (see Materials and Methods). When co-expressed, Sdk1-YFP co-clustered with the

Sdk1-RFP, but Sdk2-YFP did not, indicating that Sdk-mediated *cis*-interactions, like Sdk-mediated intercellular interactions, show homophilic specificity (Yamagata and Sanes, 2008; Hayashi et al, 2005; Figure 5F). Moreover, whereas wild-type Sdk1-YFP co-clustered with Sdk1-RFP, it did not co-cluster with Sdk1 N22R-RFP confirming that the dimer interaction is required for clustering (Figure 5F). Altogether, our data are consistent with a model whereby the crystallographically-observed Ig1–4 dimer mediates homophilic interactions in *trans* (between cells), as well as in *cis* (between Sdks on the same cell surface) (Figure 5G).

Sidekick specificity is primarily conferred by Ig1–2

Homophilic interactions of Sdk1 and Sdk2 promote specific synaptic connectivity in retina (Krishnaswamy et al., 2015; M.Y., J.R.S., and A. Krishnaswamy, in preparation). To assess the nature of this specificity in greater detail, we conducted *in vitro* SPR experiments in which we immobilized wild-type Sdk1_{Ig1–6} and Sdk2_{Ig1–6} onto an SPR chip by amine coupling, and then flowed a series of wild-type and mutant Sdk proteins over these surfaces to assess their relative binding. First we flowed wild-type Sdk1_{Ig1–6} and Sdk2_{Ig1–6} over the Sdk1_{Ig1–6} and Sdk2_{Ig1–6} surfaces. Both constructs bound to both surfaces, indicating Sdk1 and Sdk2 can bind both homophilically and heterophilically. However, homophilic binding was stronger than heterophilic binding for both the Sdk1_{Ig1–6} and Sdk2_{Ig1–6} surfaces (Figure 6A). This reveals that the segregation of Sdk1 and Sdk2 expressing cells into separate aggregates is determined not by a complete inability to interact heterophilically but rather a preference for homophilic interaction.

The results for Sdk1_{Ig1–4} and Sdk2_{Ig1–4} over the Ig1–6 SPR surfaces were comparable to those of the Ig1–6 proteins, indicating that it is the horseshoe portion of both Sdks that is responsible for both homophilic and heterophilic interaction (Figure 6A). The Ig1–2 interface mutants (Sdk1 N22R and Sdk2 H18R/N22S) showed a loss of both homophilic and heterophilic binding revealing that the heterophilic interaction observed here is mediated by the same interface as the homophilic interaction (Figure 6A). Consistent with the AUC results (Table 1), the Ig3–4 interface mutants (Sdk1 N253E and Sdk2 N253E)

showed slightly reduced binding relative to the wild-type homophilic binding. However these mutations had no impact on heterophilic binding (Figure 6A).

To further define the basis of Sdk specificity, we generated a series of chimeric constructs converting Sdk1 to Sdk2 one Ig domain at a time in the Ig1–6 context for biophysical analysis and the full length context for cell aggregation assays. However we were unable to produce these Ig1–6 chimeras at high enough solubility for accurate biophysical analysis, suggesting that the chimeras may not have fully native conformations. As noted above, a Sdk2_{Ig1–2}/Sdk1_{Ig3–4} chimera was well behaved and formed a strong dimer in solution with comparable affinity to Sdk1_{Ig1–4} and Sdk2_{Ig1–4} (Figure 3 and Table 1). Sdk2_{Ig1–2}/Sdk1_{Ig3–4} behaved more like Sdk2 than Sdk1 in the SPR assay: It bound to the Sdk1_{Ig1–6} surface similarly to Sdk2_{Ig1–4}, and more tightly to the Sdk2_{Ig1–6} surface than Sdk1_{Ig1–4} although not to the same level as Sdk2_{Ig1–4} (Figure 6A). These results suggest that the majority of the Sdk1/Sdk2 specificity is localized to Ig1–2, but that Ig3–4 also contributes to specificity.

Analysis of the crystal structures revealed that there are two Ig1–2:Ig1–2 interfacial residues that interact with one another and show conserved differences between Sdk1 and Sdk2 (Figure 4 and Figure 4—figure supplement 1). These residues—29_{Ig1} and 168_{Ig2}—therefore seem good candidates for being involved in specificity. In Sdk1, L29 interacts with the aliphatic portion of the E168 side chain, with the charged carboxylate group extending away from the hydrophobic leucine. In Sdk2 the M29 side chain interacts with D168. Since the aspartate is shorter than the glutamate in Sdk1 the charge is nearer the more polar methionine in Sdk2, which is likely better able to accommodate the charged residue than leucine (Figure 6B). To determine whether these residues play a role in specificity, the Sdk1_{Ig1–4} L29M/E168D double mutant was generated. These mutations did not impair homodimerization in solution (Table 1) or mediation of cell-cell adhesion (Figure 5B). However this mutant did affect specificity, showing weaker binding to the Sdk1_{Ig1–6} surface relative to wild-type Sdk1_{Ig1–4} and stronger binding to the Sdk2_{Ig1–6} surface consistent with these residues playing a role in specificity (Figure 6A). However the Sdk1_{Ig1–4} L29M/E168D mutant did not bind to the Sdk2_{Ig1–6} surface as strongly as the Sdk2_{Ig1–2}/Sdk1_{Ig3–4} chimera or

wild type Sdk2_{Ig1-4}, indicating that additional Ig1–2 residues also contribute to the observed specificity, alongside the Ig3–4 contribution. As already noted above, there are three additional Ig1–2 interfacial residues that show conserved differences between Sdk1 and Sdk2 (I/V17, K/R55, I/P135), which may also contribute to specificity.

Discussion

The cell adhesion molecules Sdk1 and Sdk2, like Dscams and CNTNs, provide molecular cues that determine the specificity of particular synaptic connections in the retina (Yamagata and Sanes, 2008). Here we have presented multiple crystal structures of the Sdk1 and Sdk2 Ig-domain mediated homodimers that are required for Sdk-mediated cell-cell adhesion. The dimers are formed by interaction between the Ig1–4 horseshoe regions of the Sdks. The dimers are primarily mediated by anti-parallel interactions between the Ig1–2 portions of the horseshoes, and mutagenesis studies demonstrated that this interaction was necessary for Sdk dimerization in solution and Sdk-mediated cell-cell adhesion. The Ig3–4 halves of the horseshoes, which interact in an anti-parallel manner in some but not all of the Sdk1 and Sdk2 crystal structures, only contribute a small amount to the binding energy. A previous study proposed Ig2 beta strand residues Q147–A152 were involved in the Sdk adhesive interaction (Hayashi et al, 2005), however our structures show that this strand is not located in the dimer interface. Hayashi et al (2005) sought to demonstrate the involvement of this beta strand by deletion of these residues, which would dramatically affect the fold of Ig2, explaining the loss of cell-cell adhesion they observed in cell expressing this mutant.

The horseshoe-motif mediated Sdk dimers presented here are very different from the horseshoe-motif mediated homodimers of Dscam1, neurofascin, and CNTN, for which there are published crystal structures (Figure 3–figure supplement 1). The human neurofascin and human CNTN2 horseshoe-motif dimers are both primarily mediated by anti-parallel interactions between the Ig2 G strands, with limited involvement from Ig1 and none from Ig3 and Ig4 (Figure S3–figure supplement 1A) (Liu et al., 2011;

Mortl et al., 2007). The *Drosophila* Dscam1 dimer is mediated by horseshoe-motif domains Ig2 and Ig3, alongside Ig7 (Figure 3–figure supplement 1C) (Wojtowicz et al., 2004; Sawaya et al., 2008). The horseshoe portion of this interaction involves Ig2:Ig2 and Ig3:Ig3 interactions, and is not sufficient for dimerization alone (Meijers et al., 2007), with the Ig7:Ig7 interaction also required (Wojtowicz et al., 2007; Sawaya et al., 2008). Therefore different IgSF members utilize different horseshoe-motif surfaces to mediate homophilic interactions, highlighting the versatility of the motif. Heterophilic interactions have also been reported for horseshoe-motif proteins, including binding of CNTNs 1, 3, 4, 5 and 6 to receptor protein tyrosine phosphatase gamma (PTP γ) (Bouyain and Watkins, 2010). The crystal structure of CNTN4 with the carbonic anhydrase-like domain from PTP γ shows that horseshoe domains Ig2 and Ig3 mediate this interaction, utilizing the same face for interaction as the Dscam1 homodimer (Bouyain and Watkins, 2010). To date no heterophilic interactions have been reported for the Ig domain regions of Sdk, however the opening of the Ig3–4 regions in the Sdk dimers could in principle accommodate interaction with another molecule, or potentially the FNIII portion of the Sdk extracellular region. Indeed, the flexibility of the Sdk dimer interaction that we have observed in our Sdk1 and Sdk2 crystal structures is surprising and could be explained by the need to accommodate an additional interaction.

In isolated cells immunostaining revealed Sdk1 localization to cell surface puncta, suggesting that Sdk molecules form discrete clusters. Disrupting Sdk1 Ig1–4-mediated dimerization by mutation of a key interfacial residue (N22R) resulted in a loss of this Sdk clustering in isolated cells. Instead the Sdk1 N22R mutant showed a diffuse expression across the whole cell surface (Figure 5E). This indicates that the Sdk1 Ig1–4-mediated dimer is required for Sdk clustering. For some cell surface molecules such as classical cadherins, binding *in cis* through the cell-cell recognition interface is geometrically unfavorable due to the rigidity of the interacting molecules. However since in Sdks there are 13 FNIII domains and two additional Ig domains separating the horseshoe ‘dimerization’ region from the membrane, the molecules are sufficiently flexible such that there should be no geometric constraints preventing the formation of a *cis* dimer. The observed Ig1–4-dimer-dependent *cis*-clusters are clearly larger assemblies

than dimers; therefore, additional interactions involving the intracellular, FNIII, or Ig domain regions are also likely to contribute to their formation.

When we mixed Sdk1 and Sdk2 expressing cells in our cell assays, we observed few mixed aggregates. In contrast, in solution Sdk Ig1–4 proteins dimerize both homophilically and heterophilically although the homophilic interactions are significantly stronger. We suggest three possible explanations for this apparent discrepancy. First, cell aggregation depends on the affinities of adhesion molecules; so strong homophilic interactions will lead to the formation of separate aggregates even when weaker heterophilic interactions can also occur (Katsamba et al., 2009). Second, we cannot exclude the possibility that determinants proximal to Ig1–4 enhance homophilic or attenuate heterophilic interactions. Third, and most interesting, competition between *cis* and *trans* interactions could sharpen specificity. The Sdk Ig1–4-mediated dimer is required for both Sdk *cis*-clustering in isolated cells and Sdk-mediated cell-cell adhesion. Since both interactions are mediated by the same interface, *cis* and *trans* dimers are likely to be mutually exclusive. Thus, *cis*-dimers formed between Sdks on the same cell surface would need to dissociate in order for Sdk dimers to form *in trans* between molecules on adjacent cells (Figure 7). Since homophilic interactions are stronger than heterophilic interactions, heterophilic *trans* interactions would not be able to outcompete homophilic *cis* interactions, ensuring that no adhesion will be observed between cells containing different Sdks. An analogous logic is used to enhance binding specificity in classical cadherins which dimerize through a strand swapping mechanism (Chen et al., 2005; Vendome et al., 2011; Vendome et al., 2014)

Cis-clustering and *cis/trans* competition could enhance specific connectivity in the crowded, complex neuropil of the central nervous system. For example, although most Sdk-positive retinal neurons express either Sdk1 or Sdk2, one population expresses both Sdks (Krishnaswamy et al., 2015). *Cis*-clustering could lead to formation of distinct Sdk1 and Sdk2 subdomains, which could engage different synaptic partners. Moreover, at least in retina, cells that express only one Sdk arborize in close proximity to cells

that express the other Sdk (Krishnaswamy et al., 2015). *Cis/trans* competition could play an important role in preventing the formation of inappropriate contacts between these populations.

In summary, the homophilic specificities of Sdk1 and Sdk2 play a crucial role in defining the patterning of synaptic connections in the retina. Our data suggest that these cell-patterning effects are based on the relative affinity of homophilic and heterophilic binding in the context of competition between *cis* and *trans* Sdk interactions. Further work will be necessary to demonstrate this in a neuronal context, as well as to determine the functional role of Sdk *cis* clustering.

Materials and Methods

Protein expression and purification

cDNA sequences for the Ig1–4 (residues 1–379), Ig1–5 (residues 1–476) and Ig1–6 (residues 1–569) portions of the mouse Sdk1 (NCBI NM_177879) and Sdk2 (NCBI NM_172800) extracellular regions (excluding the native signal sequence) were cloned into the mammalian expression vector p α SHP-H mammalian expression vector (a kind gift from Daniel J. Leahy, John Hopkins University), which contains the DNA sequences for the PTPa signal sequence, N-terminal hexahistidine and Strep tags, and the cleavage recognition site for Prescission protease. Point mutations were generated by the Quikchange method (Stratagene) using the KOD Hot Start DNA polymerase (Novagen). The Sdk2_{Ig1–2}/Sdk1_{Ig3–4} chimera (Sdk2 residues 1–188 and Sdk1 residues 189–379) was assembled by PCR.

Constructs were transfected into adhesive human embryonic kidney (HEK)-293 GNTi- cells (Sdk1_{Ig1–4} for crystallization) or suspension HEK-293F cells (all other proteins) (Invitrogen) with phenylethylamine (Polysciences Inc.). Conditioned media was collected after 5–6 days. The secreted Sdk proteins were purified by nickel-affinity chromatography—followed by anion and cation exchange chromatography for Sdk1_{Ig1–4} from GNTI- cells—and, after cleavage of the tag using Prescission protease (Invitrogen), size exclusion chromatography using an Akta FPLC system (GE Healthcare). The purified proteins were

concentrated to 3–8.5mg/ml in 10 mM Tris-Cl pH8, 150 mM sodium chloride (all Sdk1 proteins) or 10 mM Bis-Tris pH6, 150 mM sodium chloride (all Sdk2 proteins and the Sdk2_{Ig1-2}/Sdk1_{Ig3-4} chimera). Purified Prescission-cleaved proteins N-terminal residues are GPALA for Sdk1 and GPAGA for Sdk2, followed by the predicted mature N-termini: QDD for both Sdk1 and Sdk2. Residue numbering is from this glutamine (Q1). The mature N-termini were predicted using the SignalP 4.0 server (Petersen et al., 2011).

X-ray crystallography

Crystals were grown in 1–2μL drops using the vapor-diffusion method at 22°C. Crystallization conditions were as follows, with added cryoprotectants given in parentheses: Sdk1_{Ig1-4} crystal form 1, 12–16% (w/v) PEG4000, 0.2 M ammonium citrate (30% (v/v) ethylene glycol, with 0.2 M cesium iodide); Sdk1_{Ig1-4} crystal form 2, 10% PEG8000, 0.01 M zinc chloride, 0.1 M HEPES, pH 7.5 (30% (v/v) ethylene glycol); Sdk1_{Ig1-5}, 1.4 M ammonium sulfate, 0.5 M lithium chloride, 10 mM yttrium(III) chloride (30% (v/v) glycerol); Sdk2_{Ig1-4} crystal form 1, 1.5 M ammonium sulfate, 0.1 M sodium citrate pH 4 (30% (v/v) glycerol); Sdk2_{Ig1-4} crystal form 2, 24% PEG3350 (w/v), 0.2 M sodium chloride, 0.1 M Bis-Tris pH 5.5 (20% (v/v) ethylene glycol); Sdk2_{Ig1-4} H18R/N22S, 13.5% PEG3350 (w/v), 0.1 M ammonium sulfate, 0.2 M sodium dihydrogen phosphate (25% (v/v) glycerol); Sdk2_{Ig1-2}/Sdk1_{Ig3-4}, 3% (w/v) PEG8000, 40% (v/v) 2-methyl-2,4-pentanediol, 0.1 M sodium cacodylate pH 6.9.

Diffraction data was collected from single crystals flash frozen at 100 K on the X4A and X4C beamlines at National Synchrotron Light Source, Brookhaven National Laboratory or the Northeastern Collaborative Access Team beamlines 24-ID-C/E at the Advanced Photon Source, Argonne National Laboratory. The data for both Sdk1_{Ig1-4} crystal forms were processed using HKL2000 (Otwinowski and Minor, 1997) and scaled/merged with SCALEPACK (Otwinowski and Minor, 1997); all other data were processed with iMOSFLM (Battye et al., 2011) and scaled/merged with SCALA or AIMLESS (Evans, 2006; Evans and Murshudov, 2013). Sdk1_{Ig1-4} crystal form 1 was solved by single wavelength anomalous diffraction

(SAD) phasing method using crystals cryo-soaked with iodide ions. Substructure solution, phasing, and density modification were carried out using AUTOSOL in PHENIX (Adams et al, 2010); model building and refinement were carried out using COOT (Emsley et al., 2010) and PHENIX respectively. The other structures were solved by molecular replacement with PHASER (McCoy et al., 2007) using the Sdk1_{Ig1-4} structure as a search model. Iterative model building and refinement were carried out using COOT and PHENIX, to yield the final refined structures whose statistics are detailed in Table S1.

Structural analysis

Protein interface buried surface areas were obtained using the 'protein interfaces, surfaces and assemblies' service (PISA) from the European Bioinformatics Institute (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html; Krissinel and Henrick, 2007). Interdomain angles were calculated using UCSF Chimera (Pettersen et al., 2004). Root mean square deviations over aligned C α atoms between structures were calculated using Pymol (Schrödinger, LLC). Crystal structure figures were made using Pymol or UCSF Chimera.

Electrostatic potential calculations

Prior to the electrostatic potential calculation the Sdk1_{Ig1-5} structure and Sdk2_{Ig1-5} structure/model were prepared as follows. The Sdk2_{Ig1-4} crystal form 1 structure was used with an Sdk2 Ig5 model, which was generated with MODELLER (Webb and Sali, 2014) using Sdk1_{Ig1-5} as a template. Missing segments from both structures were also built using the MODELLER program, using the other chain or another crystal form structure as a template. Hydrogen atoms and missing side chain atoms were built with the CHARMM program (MacKerell et al., 1998). The structures were then subjected to a two-step minimization (conjugate gradient method) implemented in NAMD (Phillips et al., 2005) with the CHARMM force field (MacKerell et al., 1998). In the first minimization step, hydrogen atoms were minimized for 3,000 steps with strong harmonic constraints of 50 kcal/mol Å² applied to non-hydrogen atoms. In the second step, strong harmonic constraints were applied to the backbone atoms and to all

Asn173 atoms, which is covalently attached to NAG, while side chain atoms were minimized with constraints for 5,000 steps. The interfacial NAG was subsequently added to the minimized structures.

Electrostatic potentials were obtained by solving the Poisson–Boltzmann equation using finite difference methods as implemented in the DelPhi program (Honig and Nicholls, 1995). Atomic radii and charges were taken from CHARMM (MacKerell et al., 1998). The interior of the proteins and water were modeled as dielectric media with dielectric constants of 2 and 80, respectively. Ionic strength was set to 0.145 M and an ion exclusion radius of 2 Å was used. The numerical calculation of the potential was iterated to convergence, defined as the point at which the potential changes $<10-5 \text{ kT e}^{-1}$ between successive iterations. DelPhi calculations were run on a cubic lattice with four focusing steps of increasing resolution (from 0.5 to 2.6 grids per Å). Visualization of electrostatic surfaces was carried out with UCSF Chimera (Pettersen et al., 2004).

Sequence conservation logos

Orthologs for Sdk1 and Sdk2 were identified using NCBI BLAST (Altschul et al., 1997). The species for which orthologs were identified are listed in Table S5. Multiple sequence alignments of the Sdk1 and Sdk2 orthologs were generated using COBALT (Papadopoulos et al., 2007) and sequence logos were generated using WebLogo3 (Crooks et al., 2004).

Sedimentation equilibrium analytical ultracentrifugation measurements

Experiments were performed in a Beckman XL-A/I analytical ultracentrifuge (Beckman-Coulter, Palo Alto CA, USA), utilizing six-cell centerpieces with straight walls, 12 mm path length and sapphire windows. All proteins were dialyzed overnight and then diluted to appropriate concentration in dialysis buffer. For all Sdk1 proteins 150 mM sodium chloride, 10 mM Tris-Cl pH 8.0 were used, and for all Sdk2 proteins and the Sdk2_{Ig1-2}/Sdk1_{Ig3-4} chimera 150 mM sodium chloride, 10 mM Bis-Tris pH 6.0 were used. Samples were diluted to an absorbance at 10 mm and 280 nm of 0.65, 0.43 and 0.23 in channels A, B and

C, respectively. Dilution buffer were used as blank. All samples were run at four speeds. Ig1–5 and Ig1–6 constructs were run at 9,000, 11,000, 13,000 and 15,000 rpm, Ig1–4 constructs at 12,000, 16,000, 20,000 and 24,000 rpm. The lowest speed was held for 20 h after which four UV-scans were taken with 1 h interval, the second lowest speed held for 10 h, followed by four scans as above, the third lowest and the highest speed performed identically as the second lowest speed. All measurements were done at 25°C, and detection was by UV at 280 nm. Solvent density and protein $\bar{v}$ were determined using the program SdnTerp (Alliance Protein Laboratories, Corte Cancion, Thousand Oaks, CA, USA). Intact molecular weights of the purified Sdk proteins were obtained using matrix-assisted laser desorption/ionization (MALDI) mass spectrometry conducted by the Columbia University mass spectrometry facility. For calculation of dimeric K_d and apparent molecular weight, all useful data were used in a global fit, using the program HeteroAnalysis, obtained from University of Connecticut (www.biotech.uconn.edu/auf).

Surface plasmon resonance (SPR) binding experiments

SPR binding assays were performed using a Biacore T100 biosensor equipped with a Series S CM4 sensor chip. Mouse Sdk1_{Ig1–6} and Sdk2_{Ig1–6} were immobilized over independent flow cells using amine-coupling chemistry in HBS pH 7.4 (10 mM HEPES, 150 mM NaCl) buffer at 35°C using a flow rate of 20 μ L/min. Dextran surfaces were activated for 10 minutes using equal volumes of 0.1 M NHS (*N*-Hydroxysuccinimide) and 0.4 M EDC (1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide). Each protein was immobilized at 65 μ g/mL in 10 mM sodium acetate, pH 4.5, for 30 seconds. The immobilized surface was blocked using a 4-minute injection of 1.0 M ethanolamine, pH 8.5. Approximately 2,000 RU of each Sdk1_{Ig1–6} and Sdk2_{Ig1–6} was immobilized over each flow cell. An unmodified surface was used as a reference surface to correct for bulk refractive index shifts. Binding analysis was performed at 25°C in a running buffer of 10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.25 mg/mL BSA and 0.005% (v/v) Tween-20. Each protein was prepared in buffer at 30, 10, and 3.3 μ M (a three-fold dilution series). Each binding

cycle consisted of a 40-second association phase and a 90-second dissociation phase at a flow rate of 50 $\mu\text{L}/\text{min}$, followed by a 60-second buffer wash at 100 $\mu\text{L}/\text{min}$. Every protein concentration was tested in duplicate within the same experiment. Buffer cycles were performed before and after each concentration series to double reference the sample binding signals to correct for systematic noise and instrument drift. The data were processed using Scrubber 2.0 (BioLogic Software). Binding responses were normalized for molecular weight differences between the proteins.

Cell assay constructs and cell lines

Mouse Sdk1 and Sdk2 cDNA were cloned under pCMV promoters (Clontech, Mountain View, CA), and fused to a sequence of yellow fluorescence protein (YFP) or mCherry (RFP). We codon-optimized the mouse Sdk1 sequence since the N-terminus of mouse Sdk1 cDNA possesses a GC-rich region which was refractory to PCR. Chimeric constructs were generated by standard molecular cloning using restriction enzymes and Q5 DNA polymerase-assisted PCR (NEB, Ipswich, MA), or the Gibson assembly kit (SGI-DNA, La Jolla, CA). Mutants were generated using QuikChange site-directed mutagenesis kits (Agilent Technologies, Santa Clara, CA). A construct for overexpressing a Rapsyn::RFPnanobody fusion protein (Rapsyn::RFPnb) was generated by synthesizing a construct encoding the 90 amino acid self-association domain of Rapsyn (Ramarao and Cohen, 1998) and a mCherry-binding nanobody LaM-4 (Fridy et al, 2014) in a pCMV backbone.

Cells were cultured in DMEM supplemented with 10% fetal calf serum and penicillin/streptomycin. To generate L cell (ATCC, Manassas, VA) lines stably expressing mouse Sdk1, Sdk2, their mutants, and fluorescent protein derivatives, the sequences were cloned into a piggyBac transposon vector pXL-CAG-Zeocin-3x F2A (Martell et al., 2016). L cells were transfected with the appropriate Sdk construct together with a piggyBac transposase vector pCAG-mPBor f (Yamagata and Sanes, 2012a) using DMRIE-C (Invitrogen, Carlsbad, CA), trypsinized after 2 days, replated into larger plates, and selected with 1 mg/mL Zeocin (Invivogen, San Diego, CA) for 2–3 weeks. Surviving colonies were transferred to new

plates and screened with antibodies against Sdk or fluorescence to select clones with high and homogeneous expression. HEK-293T cells were obtained from ATCC. HEK-293T cells endogenously express human N-cadherin which results in some background in cell adhesion assays. To decrease this background, expression of N-cadherin was fully eliminated by disrupting both alleles of the N-cadherin gene (CDH2) by CRISPR toward the N-cadherin gene. Full characterization of this N-cadherin-deficient HEK-293 cell line will be described elsewhere (M.Y. and J.R.S. unpublished). HEK-293T cells were transfected with DMR1E-C in OptiMEM (Invitrogen), and used for experiments 2–3 days after transfection. For co-transfection into 293T cells, two Sdk plasmids and a Rapsyn::RFPnb plasmid were mixed in a 5:5:1 ratio.

Antibodies and immunostaining

Production of affinity-purified rabbit polyclonal antibodies to the cytoplasmic domain of mouse Sdk1 was described previously (Krishnaswamy et al. 2015). A mouse monoclonal antibody to mouse Sdk1 was generated from a Sdk1-knockout mouse (Krishnaswamy et al. 2015) immunized with mouse Sdk1-expressing L cells. Briefly, splenocytes from a hyperimmunized mouse were fused to a myeloma cell line FOXPY (ATCC), selected, and screened by immunostaining Sdk1-transfected L cells. One established hybridoma line, MS1-7, produces an IgG1/kappa monoclonal antibody that recognizes FNIII domains of mouse Sdk1 (not shown). Chicken anti-GFP antibodies were described previously (Yamagata and Sanes, 2012b). Species-specific Alexa dye-conjugated secondary antibodies were obtained from Jackson ImmunoResearch (West Grove, PA). Fluorescein-conjugated wheat germ agglutinin (WGA) was from Vector Laboratories (Burlingame, CA). Cultured cells on glass coverslips (Bellco Glass, Vineland, NJ) were fixed with 4% paraformaldehyde/PBS for 30 min at 4 °C, and then either observed directly after mounting with Fluorogel (Electron Microscopy Sciences, Hatfield, PA), or immunostained after treatment with 0.1% (w/v) TritonX-100/PBS for 10 min at room temperature. The standard immunostaining procedure used was described previously (Yamagata and Sanes, 2012a).

Cell aggregation assay

Confluent stably-transfected L cells or transiently-transfected 293T cells were trypsinized in the presence of 1 mM EDTA at 37 °C as described previously (Yamagata and Sanes, 2008). In some experiments, cells were labeled with green or red Cell Trackers (Invitrogen). The reaction was stopped by adding the same volume of 0.1 mg/ml soybean trypsin inhibitor (T6522, Sigma, St. Louis, MO) and 10 µg/ml deoxyribonuclease I (DN25, Sigma) in HBSS supplemented with 20mM HEPES, pH7.4. All the cell aggregation assays were carried out in 24-well non-tissue culture plasticwares that had been precoated with 0.5% BSA/HBSS. In each well, dissociated cells were mixed with 1 ml of HBSS containing 0.5% (w/v) BSA, 1 µg/ml deoxyribonuclease I, 20 mM HEPES, pH7.4, and rotated at room temperature for 30–60 min. The reaction was stopped by adding 1 ml of 4% (w/v) paraformaldehyde/PBS, and observed under fluorescent microscopes.

Accession Numbers

The atomic coordinates and structure factors for the reported crystal structures are deposited in the Protein Data Bank under accession codes PDB: 5K6U, 5K6V, 5K6W, 5K6X, 5K6Y, 5K6Z, and 5K70.

Author Contributions

K.M.G., M.Y., X.J., B.H., J.S., and L.S. designed research. X.J. determined the Sdk1_{Ig1-4} crystal structures. K.M.G. determined all other crystal structures. S.M. and K.M.G. produced all proteins. M.Y. conducted and analyzed all cell-based assays. G.A. performed and analyzed the AUC experiments. P.S.K. performed and analyzed the SPR experiments. A.P.S. performed bioinformatics analysis of horseshoe-motif proteins. K.M.G., M.Y., B.H., J.S., and L.S. analyzed data and wrote the manuscript.

Acknowledgements

We thank John Schwanof and Randy Abramowitz at Brookhaven National Lab XSLs X4A/C beamlines and Igor Kourinov, Surajit Banerjee, and Narayanasami Sukumar at NE-CAT Advanced Photon Source 24-ID-C/E beamlines, Argonne National Laboratory (NIH grant P41 GM103403) for their support with synchrotron data collection. We thank Rotem Rubinstein (Columbia) for the electrostatics calculations.

Competing Interests

The authors declare no competing interests.

References

- Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst. D* 66, 213–221. doi: 10.1107/S0907444909052925
- Agarwala, K.L., Ganesh, S., Amano, K., Suzuki, T. and Yamakawa, K. (2001). DSCAM, a highly conserved gene in mammals, expressed in differentiating mouse brain. *Biochemical and biophysical research communications* 281, 697–705. doi:10.1006/bbrc.2001.4420
- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R., and Leslie, A.G.W. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Cryst. D* 67, 271–281. doi: 10.1107/S0907444910048675
- Bouyain, S., and Watkins, D.J. (2010). The protein tyrosine phosphatases PTPRZ and PTPRG bind to distinct members of the contactin family of neural recognition molecules. *Proc. Natl. Acad. Sci. USA* 107, 2443–2448. doi: 10.1073/pnas.0911235107
- Chen, C.P., Posy, S., Ben-Shaul, A., Shapiro, L., and Honig, B.H. (2005). Specificity of cell-cell adhesion by classical cadherins: Critical role for low-affinity dimerization through beta-strand swapping. *Proc. Natl. Acad. Sci. USA* 102, 8531–6. doi: 10.1073/pnas.0503319102
- Chen, Q., Sun, X., Zhou, X.H., Liu, J.H., Wu, J., Zhang, Y., and Wang, J.H. (2013). N-terminal horseshoe conformation of DCC is functionally required for axon guidance and might be shared by other neural receptors. *J Cell Sci.* 126, 186–195. doi: 10.1242/jcs.111278
- Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner S.E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Cryst. D* 66, 486–501. doi: 10.1101/gr.849004
- Evans P.R. (2006). Scaling and assessment of data quality. *Acta Cryst. D* 62, 72–82. doi: 10.1107/S0907444905036693

- Evans, P. R., & Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Cryst. D* 69, 1204–1214. doi: 10.1107/S0907444913000061
- Freigang, J., Proba, K., Leder, L., Diederichs, K., Sonderegger, P., and Welte, W. (2000). The crystal structure of the ligand binding module of axonin-1/TAG-1 suggests a zipper mechanism for neural cell adhesion. *Cell* 101, 425–433. doi:10.1016/S0092-8674(00)80852-1
- Fridy, P.C., Li, Y., Keegan, S., Thompson, M.K., Nudelman, I., Scheid, J.F., Oeffinger, M., Nussenzweig, M.C., Fenyő, D., Chait, B.T., and Rout, M.P. (2014). A robust pipeline for rapid production of versatile nanobody repertoires. *Nat. Methods* 11, 1253–1260. doi: 10.1038/nmeth.3170
- Fuerst, P.G., Koizumi, A., Masland, R.H., and Burgess, R.W. (2008). Neurite arborization and mosaic spacing in the mouse retina require DSCAM. *Nature* 451, 470–474. doi: 10.1038/nature06514
- Fuerst, P.G., Bruce, F., Tian, M., Wei, W., Elstrott, J., Feller, M.B., Erskine, L., Singer, J.H., and Burgess, R.W. (2009). DSCAM and DSCAML1 function in self-avoidance in multiple cell types in the developing mouse retina. *Neuron* 64, 484–497. doi: 10.1016/j.neuron.2009.09.027
- Gouveia R. M., Gomes C. M., Sousa M., Alves P. M., and Costa J. (2008). Kinetic analysis of L1 homophilic interaction: role of the first four immunoglobulin domains and implications on binding mechanism. *J. Biol. Chem.* 283, 28038–28047. doi: 10.1074/jbc.M804991200
- Harrison, O.J., Jin, X., Hong, S., Bahna, F., Ahlsen, G., Brasch, J., Wu, Y., Vendome, J., Felsovalyi, K., Hampton, C.M., *et al.* (2011). The extracellular architecture of adherens junctions revealed by crystal structures of type I cadherins. *Structure* 19, 244–256. doi: 10.1016/j.str.2010.11.016
- Haspel, J., Friedlander, D.R., Ivgy-May, N., Chickramane, S., Roonprapunt, C., Chen, S., Schachner, M. and Grumet, M. (2000). Critical and optimal Ig domains for promotion of neurite outgrowth by L1/Ng-CAM. *J. Neurobiol.* 42, 287–302. doi: 10.1002/(SICI)1097-4695(20000215)42:3<287::AID-NEU1>3.0.CO;2-X
- Hayashi, K., Kaufman, L., Ross, M.D., and Klotman, P.E. (2005). Definition of the critical domains required for homophilic targeting of mouse sidekick molecules. *FASEB J.* 19, 614–616. doi: 10.1096/fj.04-2947fje
- Honig, B., and Nicholls, A. (1995). Classical electrostatics in biology and chemistry. *Science* 268, 1144–1149. doi: 10.1126/science.7761829
- Katsamba, P., Carroll, K., Ahlsen, G., Bahna, F., Vendome, J., Posy, S., Rajebhosale, M., Price, S., Jessell, T.M., Ben-Shaul, A., Shapiro, L., and Honig, B.H. (2009). Linking molecular affinity and cellular specificity in cadherin-mediated adhesion. *Proc. Natl. Acad. Sci. USA* 106, 11594–9. doi: 10.1073/pnas.0905349106
- Kaufman, L., Hayashi, K., Ross, M.J., Ross, M.D., and Klotman, P.E. (2004). Sidekick-1 is upregulated in glomeruli in HIV-associated nephropathy. *J. Am. Soc. Nephrol.* 15, 1721–1730. doi: 10.1097/01.ASN.0000128975.28958.C2
- Kaufman, L., Yang, G., Hayashi, K., Ashby, J.R., Huang, L., Ross, M.J., Klotman, M.E., and Klotman, P.E. (2007). The homophilic adhesion molecule sidekick-1 contributes to augmented podocyte aggregation in HIV-associated nephropathy. *FASEB J.* 21, 1367–1375. doi: 10.1096/fj.06-7191com
- Kaufman, L., Potla, U., Coleman, S., Dikiy, S., Hata, Y., Kurihara, H., He, J.C., D’Agati, V.D., and Klotman, P.E. (2010). Up-regulation of the Homophilic Adhesion Molecule Sidekick-1 in Podocytes Contributes to Glomerulosclerosis. *J. Biol. Chem.* 285, 25677–25685. doi: 10.1074/jbc.M110.133959
- Krishnaswamy, A., Yamagata, M., Duan, X., Hong, Y.K., and Sanes, J.R. (2015). Sidekick 2 directs formation of a retinal circuit that detects differential motion. *Nature* 524, 466–470. doi: 10.1038/nature14682

- Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372, 774–797. doi:10.1016/j.jmb.2007.05.022
- Liu, H., Focia, P.J., and He, X. (2011). Homophilic adhesion mechanism of neurofascin, a member of the L1 family of neural cell adhesion molecules. *J. Biol. Chem.* 286, 797–805. doi: 10.1074/jbc.M110.180281
- MacKerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., et al. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102, 3586–3616. doi: 10.1021/jp973084f
- Maness P. F., and Schachner M. (2007). Neural recognition molecules of the immunoglobulin superfamily: signaling transducers of axon guidance and neuronal migration. *Nat. Neurosci.* 10, 19–26. doi:10.1038/nn1827
- Martell, J.D., Yamagata, M., Deerinck, T.J., Phan, S., Kwa, C.G., Ellisman, M.H., Sanes, J.R., and Ting, A.Y. (2016). A split horseradish peroxidase for the detection of intercellular protein–protein interactions and sensitive visualization of synapse. *Nature Biotech.* *in press*. doi:10.1038/nbt.3563
- Masland, R.H. (2012). The neuronal organization of the retina. *Neuron* 76, 266–280. doi: 10.1016/j.neuron.2012.10.002
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. *Journal of applied crystallography* 40, 658–674. doi: 10.1107/S0021889807021206
- Meijers, R., Puettmann-Holgado, R., Skiniotis, G., Liu, J.H., Walz, T., Wang, J.H., and Schmucker, D. (2007). Structural basis of Dscam isoform specificity. *Nature* 449, 487–491. doi:10.1038/nature06147
- Mörtl, M., Sonderegger, P., Diederichs, K., and Welte, W. (2007). The crystal structure of the ligand-binding module of human TAG-1 suggests a new mode of homophilic interaction. *Protein Sci.* 16, 2174–2183. doi: 10.1110/ps.072802707
- Nguyen, D.N., Liu, Y., Litsky, M.L. and Reinke, R. (1997). The sidekick gene, a member of the immunoglobulin superfamily, is required for pattern formation in the *Drosophila* eye. *Development* 124, 3303–3312.
- Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol.* 276, 307–326. doi:10.1016/s0076-6879(97)76066-x
- Papadopoulos, J.S., and Agarwala, R. (2007). COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* 23, 1073–1079. doi: 10.1093/bioinformatics/btm076
- Petersen, T.N., Brunak S., von Heijne G., Nielsen H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi: 10.1038/nmeth.1701
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612. doi: 10.1002/jcc.20084
- Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L., and Schulten, K. (2005). Scalable molecular dynamics with NAMD. *J Comput Chem* 26, 1781–1802. doi: 10.1002/jcc.20289
- Ramarao, M.K., and Cohen, J.B. (1998). Mechanism of nicotinic acetylcholine receptor cluster formation by rapsyn. *Proc. Natl. Acad. Sci. USA* 95, 4007–4012. <http://www.jstor.org/stable/44600>.
- Webb, B., and Sali, A. (2014). Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Bioinformatics* 47, 5 6 1–32. doi: 10.1002/cpbi.3

- Roska, B., and Werblin, F. (2001). Vertical interactions across ten parallel, stacked representations in the mammalian retina. *Nature* 410, 583–587. doi:10.1038/35069068
- Sanes, J.R., and Masland, R.H. (2015). The Types of Retinal Ganglion Cells: Current Status and Implications for Neuronal Classification. *Ann. Rev. Neurosci.* 38,221–246. doi: 10.1146/annurev-neuro-071714-034120
- Sanes, J.R., and Zipursky, S.L. (2010). Design principles of insect and vertebrate visual systems. *Neuron* 66, 15–36. doi: 10.1016/j.neuron.2010.01.018
- Sawaya, M.R., Wojtowicz, W.M., Andre, I., Qian, B., Wu, W., Baker, D., Eisenberg, D., and Zipursky, S.L. (2008). A double S shape provides the structural basis for the extraordinary binding specificity of Dscam isoforms. *Cell* 119, 1007–1018. doi: 10.1016/j.cell.2008.07.042
- Schmucker, D., and Chen, B. (2009). Dscam and DSCAM: complex genes in simple animals, complex animals yet simple genes. *Genes Dev.* 23,147–156. doi: 10.1101/gad.1752909
- Shimoda, Y. and Watanabe, K. (2009). Contactins: emerging key roles in the development and function of the nervous system. *Cell adhesion & migration* 3, 64–70. doi: 10.4161/cam.3.1.7764
- Su, X.D., Gastinel, L.N., Vaughn, D.E., Faye, I., Poon, P., and Bjorkman, P.J. (1998). Crystal structure of hemolin: a horseshoe shape with implications for homophilic adhesion. *Science* 281, 991–5. doi: 10.1126/science.281.5379.991
- Vendome, J., Posy, S., Jin, X., Bahna, F., Ahlsen, G., Shapiro, L., and Honig, B. (2011) Molecular design principles underlying β -strand swapping in the adhesive dimerization of cadherins. *Nat. Struct. Mol. Biol.* 18, 693–700. doi: 10.1038/nsmb.2051
- Vendome, J., Felsovalyi, K., Song, H., Yang, Z., Jin, X., Brasch, J., Harrison, O.J., Ahlsen, G., Bahna, F., Kaczynska, A., Katsamba, P.S., Edmond, D., Hubbell, W.L., Shapiro, L., and Honig, B. (2014). Structural and energetic determinants of adhesive binding specificity in type I cadherins. *Proc. Natl. Acad. Sci. USA* 111, E4175–84. doi: 10.1073/pnas.1416737111
- Wojtowicz, W.M., Flanagan, J.J., Millard, S.S., Zipursky, S.L., and Clemens, J.C. (2004). Alternative splicing of *Drosophila* Dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding. *Cell* 118, 619–633. doi:10.1016/j.cell.2004.08.021
- Wojtowicz, W.M., Wu, W., Andre, I., Qian, B., Baker, D., and Zipursky, S.L. (2007). A vast repertoire of Dscam binding specificities arises from modular interactions of variable Ig domains. *Cell* 130, 1134–1145. doi: 10.1016/j.cell.2007.08.026
- Wu, W., Ahlsen, G., Baker, D., Shapiro, L., and Zipursky, S.L. (2012). Complementary chimeric isoforms reveal Dscam1 binding specificity in vivo. *Neuron* 74, 261–268. doi: 10.1016/j.neuron.2012.02.029
- Yamagata, M., Weiner, J. A., and Sanes, J. R. (2002). Sidekicks: synaptic adhesion molecules that promote lamina-specific connectivity in the retina. *Cell* 110, 649–660. doi:10.1016/S0092-8674(02)00910-8
- Yamagata, M., and Sanes, J. R. (2008). Dscam and Sidekick proteins direct lamina-specific synaptic connections in vertebrate retina. *Nature* 451, 465–469. doi: 10.1038/nature06469
- Yamagata, M., and Sanes, J.R. (2010). Synaptic localization and function of Sidekick recognition molecules require MAGI scaffolding proteins. *J. Neurosci.* 30, 3579–3588. doi: 10.1523/JNEUROSCI.6319-09.2010
- Yamagata, M., and Sanes, J.R. (2012a). Expanding the Ig superfamily code for laminar specificity in retina: expression and role of contactins. *J. Neurosci.* 32, 14402–14414. doi: 10.1523/JNEUROSCI.3193-12.2012

Yamagata M., and Sanes, J.R. (2012b). Transgenic strategy for identifying synaptic connections in mice by fluorescence complementation (GRASP). *Front Mol. Neurosci.* 5, 18. doi: 10.3389/fnmol.2012.00018

Yamakawa, K., Huot, Y.K., Haendelt, M.A., Hubert, R., Chen, X.N., Lyons, G.E., and Korenberg, J.R. (1998). DSCAM: a novel member of the immunoglobulin superfamily maps in a Down syndrome region and is involved in the development of the nervous system. *Hum. Mol. Genet.* 7, 227–237. doi: 10.1093/hmg/7.2.227

Table

Protein	Oligomeric state	Dimerization K_D (μM , $n=2$)
<u>Sdk1</u>		
Ig1-4	Dimer	10.5 ± 1.1
Ig1-5	Dimer	4.6 ± 0.06
Ig1-6	Dimer	2.3 ± 0.39
<i><u>Ig1-2 dimer interface mutations</u></i>		
Ig1-4 N22R	Monomer	N/A
Ig1-4 K133E	Weak non-specific dimer	204 ± 38.9 ($K_I/K_D=1.24$)
Ig1-4 L29M/E168D	Dimer	4.26 ± 0.50
Ig1-6 N22R	Monomer	650 ± 66 ($K_I/K_D=1.05$)
<i><u>Ig3-4 dimer interface mutation</u></i>		
Ig1-4 N253E	Dimer	15.5 ± 1.64
<u>Sdk2</u>		
Ig1-4	Dimer	2.2 ± 0.4
Ig1-5	Dimer	0.73 ± 0.036
Ig1-6	Dimer	0.44 ± 0.012
<i><u>Ig1-2 dimer interface mutations</u></i>		
Ig1-4 H18R/N22S	Monomer	N/A
Ig1-6 H18R/N22S	Monomer	N/A
<i><u>Ig3-4 dimer interface mutation</u></i>		
Ig1-4 N253E	Dimer	18.9 ± 0.95
<u>Chimera</u>		
Sdk2 _{Ig1-2} /Sdk1 _{Ig3-4}	Dimer	3.92 ± 0.17

Table 1: Sedimentation equilibrium analytical ultracentrifugation of Sdk fragments

K_D = dissociation constant. K_I = isodesmic constant. The K_I/K_D ratio is given when it is less than two, indicating the presence of non-specific binding.

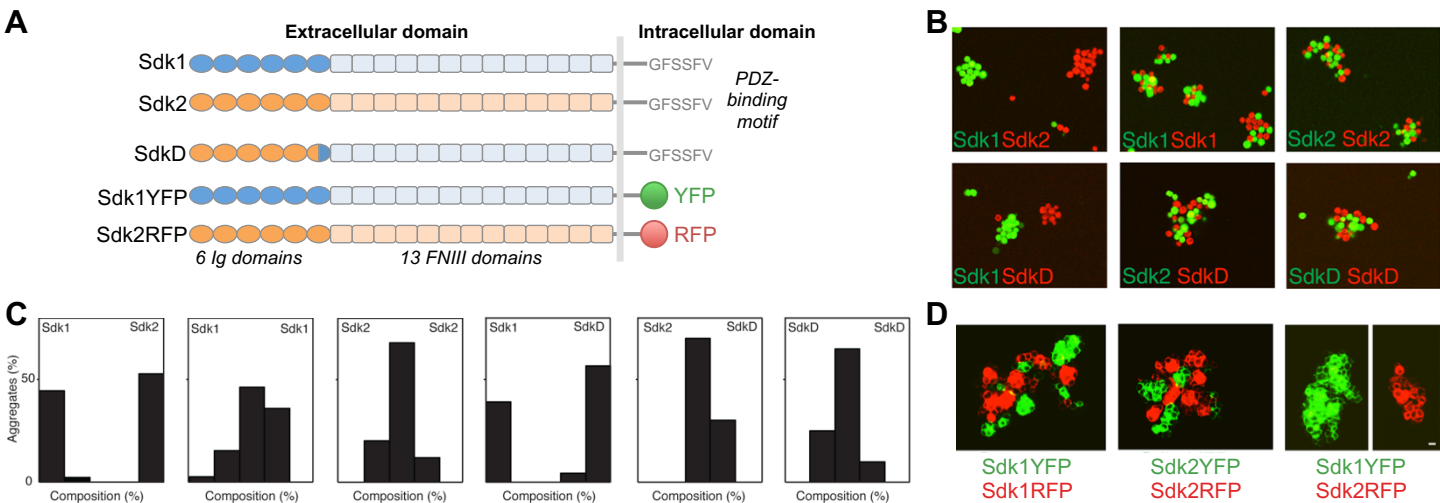


Figure 1. The Sdks' Ig domain regions are responsible for Sdk1/2 specificity.

A. Schematic of the domain arrangement of mouse Sdk1, Sdk2, and Sdk constructs used in the cell assays.

B. Aggregation assays demonstrate that L cells (labeled with red or green cell-trackers) co-aggregate only when the cells are expressing the same Sdk (top row). Cells expressing the SdkD chimera (shown in **A**) co-aggregate with Sdk2 but not Sdk1 expressing cells (bottom row). Scale bar, 5 μ m.

C. Quantification of the assays shown in **B**. Each graph shows the proportion of red and green cells observed in the cell aggregates in each assay (n=20–39).

D. HEK-293 cells transiently transfected with Sdk-RFP and YFP constructs show cell surface fluorescence and only co-aggregate with cells expressing the same Sdk molecule, regardless of the intracellular fluorescent tag. Scale bar, 5 μ m.

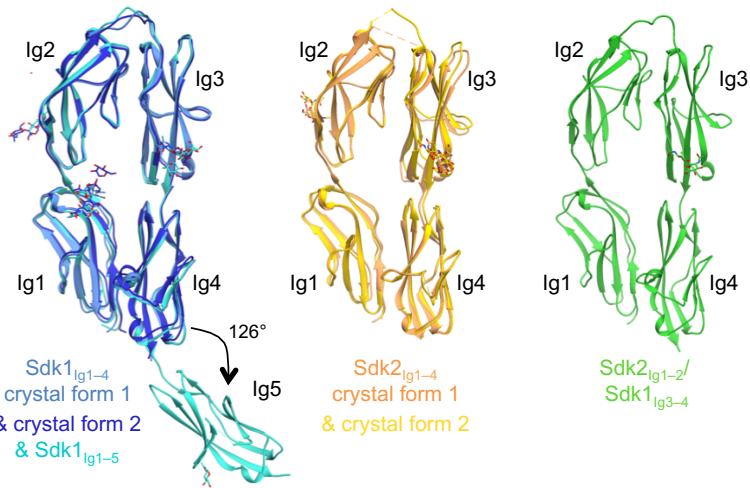
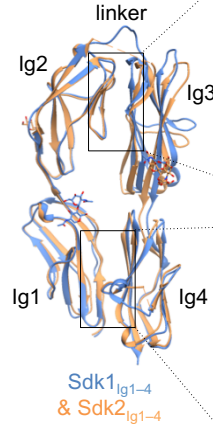
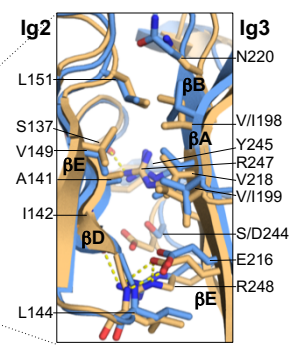
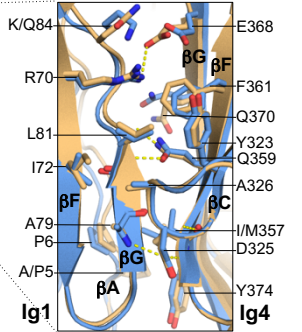
A**B****C****D**

Figure 2. The N-terminal four Ig domains of Sdk1 and Sdk2 adopt a stable horseshoe conformation.

A. Crystal structures of the five or four N-terminal Ig domains of Sdk1 and Sdk2 show a horseshoe arrangement of Ig1–4, which is maintained in a Sdk2_{Ig1-2}/Sdk1_{Ig3-4} chimera, from which Ig5 extends at a 126° angle from Ig4 in Sdk1. The left hand panel shows a structural alignment of single protomers from the two different Sdk1_{Ig1-4} crystal forms and the Sdk1_{Ig1-5} structure, showing the very high structural similarity of individual protomers among the different crystal forms. The second panel shows protomers from the two Sdk2_{Ig1-4} crystal forms structurally aligned, and the third panel shows a single protomer from the Sdk2_{Ig1-2}/Sdk1_{Ig3-4} structure. All protein structures are shown in ribbon depiction, with oligosaccharides in stick view.

B. Structural alignment of the Sdk1 and Sdk2 Ig1–4 structures revealing that their overall architecture is highly similar.

C. Close-up of the Sdk1 (blue) and Sdk2 (orange) Ig2:Ig3 intra-horseshoe interfaces with interfacial residue side-chains and key main-chain atoms shown as sticks. Sdk2 hydrogen bonds are depicted as dashed yellow lines.

D. Close-up of the Ig1:Ig4 intra-horseshoe interface.

See also Figure 2–supplemental figure 1 and Tables S1, S2, and S3.

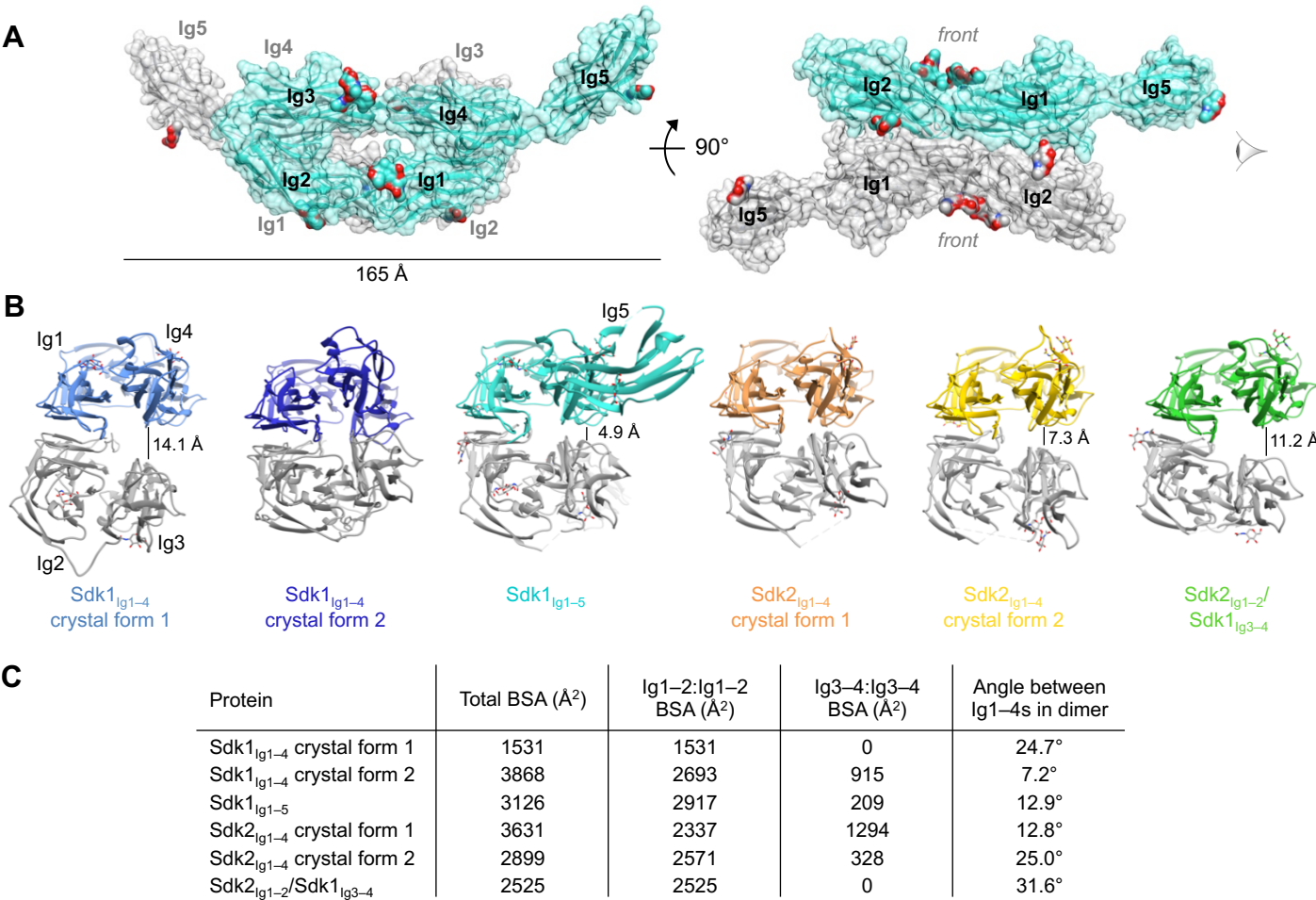


Figure 3. Crystal structures of Sdk1 and Sdk2 reveal a flexible dimer of horseshoes mediated by the N-terminal four Ig domains.

A. Surface view of the Sdk1_{Ig1-5} structure showing the back-to-back horseshoe-mediated dimer observed in the crystal structure. One protomer is colored cyan, one protomer colored gray. Oligosaccharides are shown as opaque surfaces, colored by element.

B. The same back-to-back dimer is observed in all Sdk structures, but the angle between the two protomers in each dimer varies. Dimer structures are shown in ribbon depiction, viewed from the eye icon in **E**. In each dimer, one protomer is colored variably and one protomer is colored gray. All variably colored protomers are aligned, so that the difference in angle is evident from the differing positions of the gray protomer.

C. Table detailing the variation amongst the Sdk dimer crystal structures in the buried surface area (BSA) over both protomers, and in the angle between the planes of the two horseshoe regions (Ig1-4) in the dimers. See also Figure 3—supplemental figure 1.

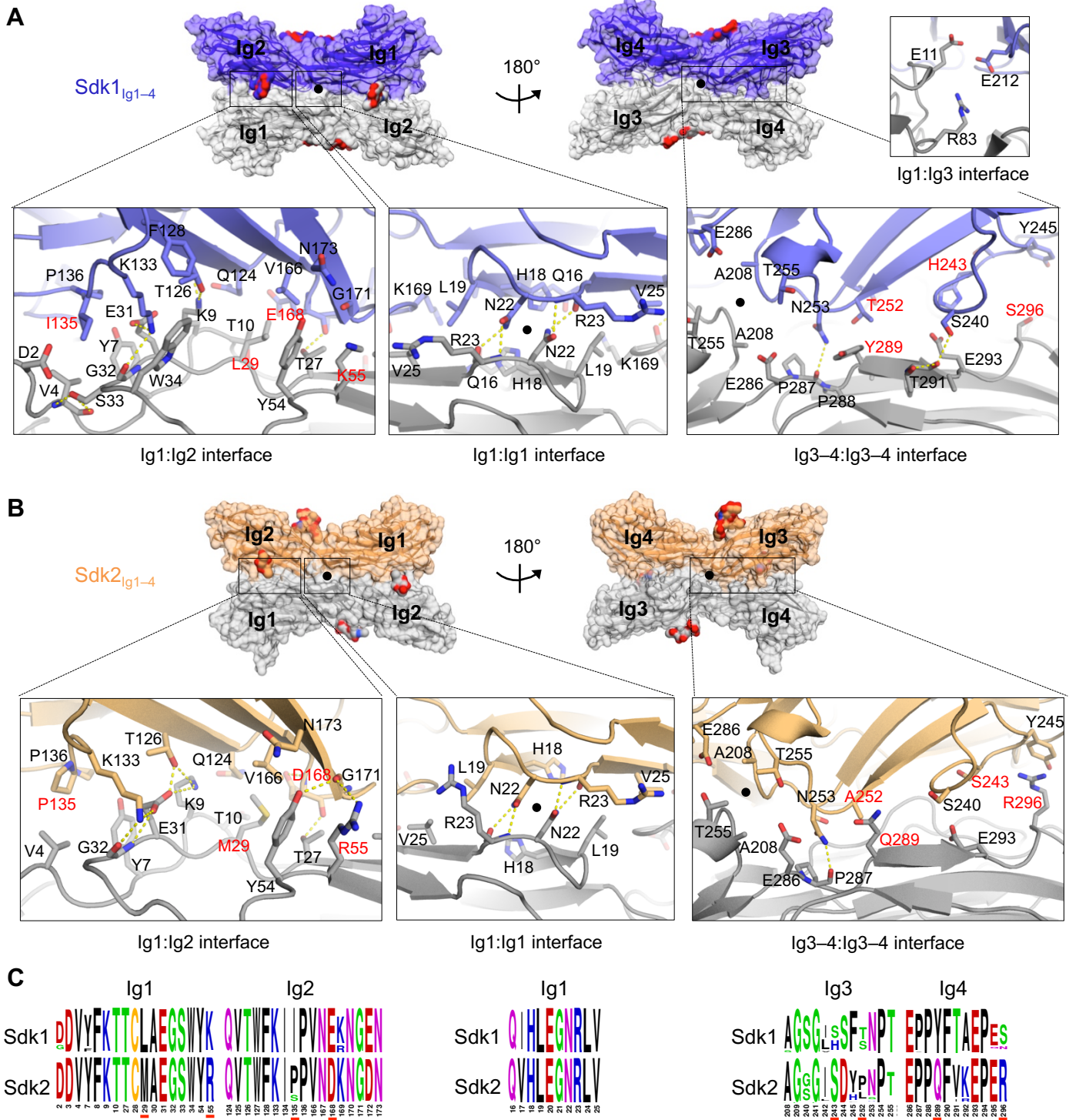


Figure 4. The Sdk1 and Sdk2 dimer interfaces are highly similar.

A. Sdk1 dimer interface. Top panel: Sdk1_{Ig1-4} crystal form 2 dimer shown in surface view, one protomer dark blue, the other gray. Bottom panel: Close ups of the Ig1:Ig2, Ig1:Ig1, Ig3-4:Ig3-4 interfaces that make up the dimer. All interfacial residue side-chains, and main-chain atoms involved in hydrogen bonds, are shown in stick view. Residues labeled in red differ between Sdk1 and Sdk2. The small Ig1:Ig3 interface observed only in the Sdk1_{Ig1-4} crystal form 2 is shown in the top right.

B. Sdk2 dimer interface. Sdk2_{Ig1-4} crystal form 1 dimer shown as for Sdk1 in **A**. One protomer is colored orange, the other gray.

C. Sequence logos of Sdk1 and Sdk2 dimer interface residues generated from multiple sequence alignments of vertebrate Sdk1 and Sdk2 protein sequences (species used are listed in Table S5).

See also Figure 4-supplemental figures 1-3, and Tables S4 and S5.

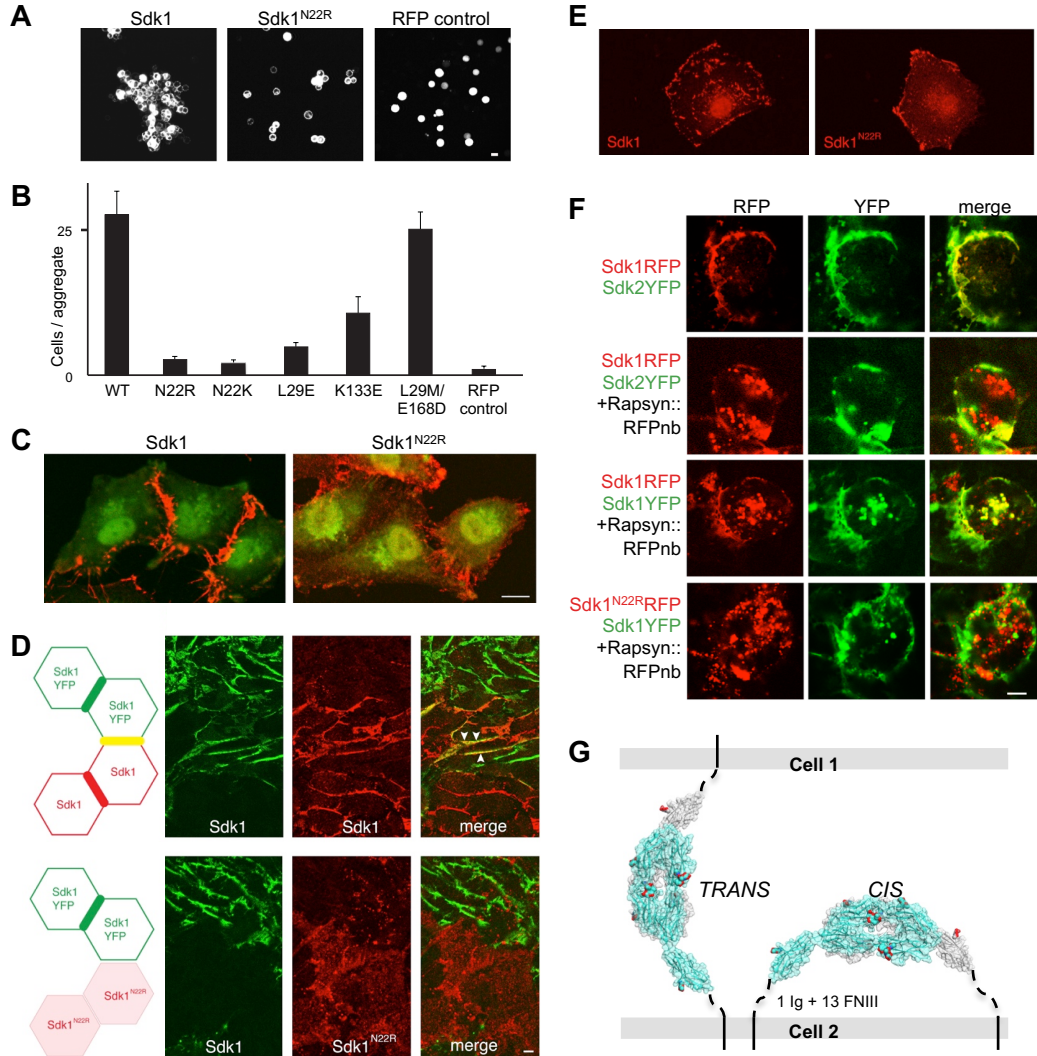


Figure 5. Sdk1 dimerization is required for cell aggregation and cis clustering.

A. Aggregation assay using N-cadherin deficient HEK-293 cells transiently transfected with wild-type and mutant Sdk1-RFP. Cytoplasmic RFP transfection was used as a negative control (right panel).

B. Quantification of the aggregation assay shown in **A** for wild-type Sdk1 and several Sdk1 dimer interface mutants (n=15, mean±S.E.).

C. Immunolabeling of Sdk1 (red) with a monoclonal antibody to Sdk1 in interacting L cells shows wild-type Sdk1 (left panel) localizes to the cell-cell junctions whereas the Sdk1 N22R mutant is diffusely localized (right panel). Counterstaining with wheat germ agglutinin (WGA, green) was used to visualize the cell surface.

D. Immunolabeled Sdk1 (red, stained with anti-Sdk1 cytoplasmic domain) and Sdk1YFP (green, Sdk1's cytoplasmic domain was replaced with YFP) co-localize at cell-cell junctions between Sdk1 and Sdk1YFP expressing L cells (top panels). Arrows indicate co-localization of red and green fluorescence. By contrast, Sdk1 N22R does not localize to cell-cell junctions between Sdk1 N22R and Sdk1YFP expressing cells (bottom panels).

E. Immunolabeling of Sdk1 with a monoclonal antibody to Sdk1 (red) in solitary L cells shows that wild-type Sdk1 localizes in puncta on the cell surface (top panels) whereas the Sdk1 N22R mutant is diffusely localized (bottom panels).

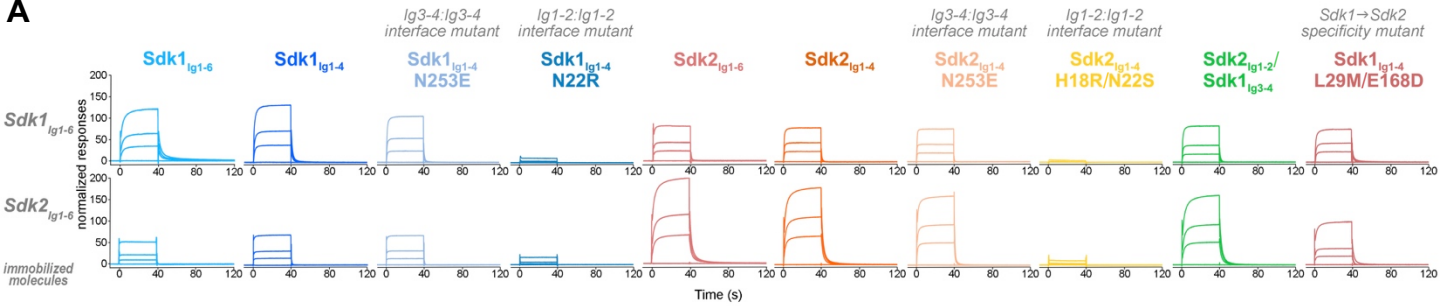
F. HEK-293 cells transiently transfected with both Sdk1RFP and Sdk1YFP express both proteins, which co-localize to cell-cell junctions (top row). Co-transfection of a Rapsyn::RFPnanobody induces clustering of Sdk1RFP away from cell-cell junctions. Sdk2YFP does not co-cluster with Sdk1RFP (second row), but Sdk1YFP does co-cluster (third row). However Sdk1YFP does not co-cluster with Sdk1 N22R-RFP/Rapsyn::RFPnanobody clusters (bottom row).

G. Our data suggest Sdk dimerizes using the crystallographically-determined interface, both between molecules emanating from opposing cell surfaces (in *trans*)—mediating cell-cell interactions—and between molecules emanating from the same cell surface (in *cis*)—mediating Sdk clustering. These interactions are shown schematically, using the Sdk1_{Ig1–5} crystal structure to illustrate the dimer interaction. The remaining 1 Ig and 13 FNIII domains that constitute the rest of the Sdk extracellular domain are abbreviated to a dashed line, with the transmembrane and intracellular domains shown as solid lines. Scale bars in **A**, **C**, **D**, **E** and **F**, 5 μm.

See also Figure 5—supplemental figure 1.

Figure 6

A



B

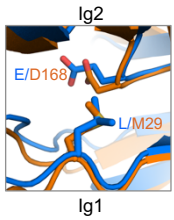


Figure 6. Sdk1 and Sdk2s' homophilic interactions are stronger than their heterophilic interaction

A. Heterophilic and homophilic Sdk binding were analyzed by a surface plasmon resonance (SPR) experiment. Sdk1_{Ig1-6} (top row) and Sdk2_{Ig1-6} (bottom row) were covalently attached to the SPR chip and three different identical concentrations of each Sdk analyte (columns) were flowed over the Sdk1_{Ig1-6} and Sdk2_{Ig1-6} surfaces. The binding association and dissociation are shown by the normalized SPR response.

B. Close up of interacting specificity residues 29 and 168 in the Sdk1 (blue) and Sdk2 (orange) dimer interfaces.

Figure 7

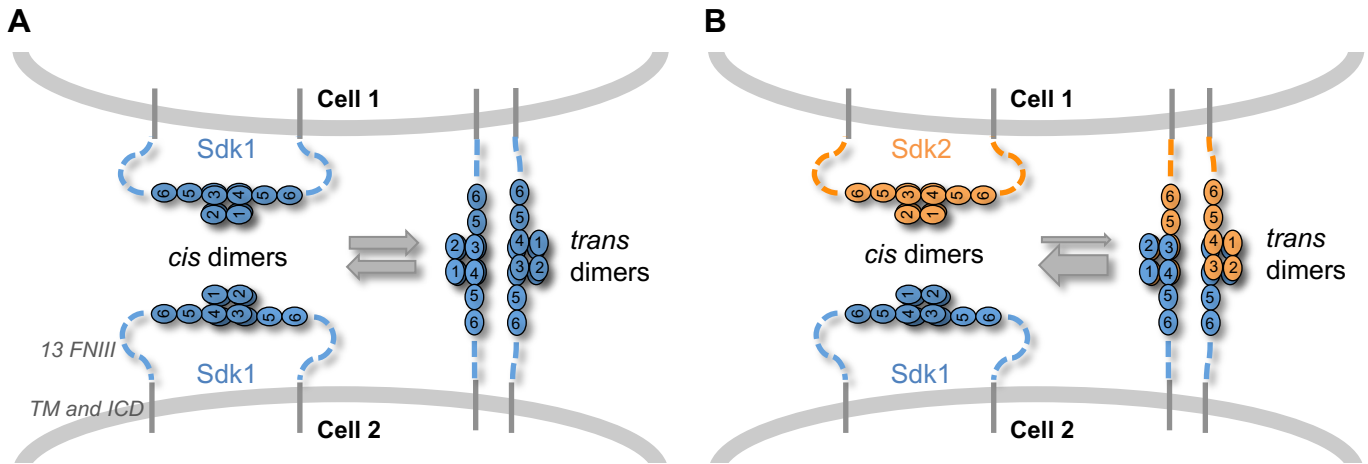


Figure 7. Sdk specificity could be accentuated by competition between *cis* and *trans* interactions.

A. Schematic of two neighboring cells expressing Sdk1 showing the competition between homophilic dimer interactions *in cis* and *in trans*.

B. Schematic of two neighboring cells, one expressing Sdk1 and one expressing Sdk2, showing the competition between the stronger homophilic dimer interactions *in cis* and the weaker heterophilic dimer interactions *in trans*.

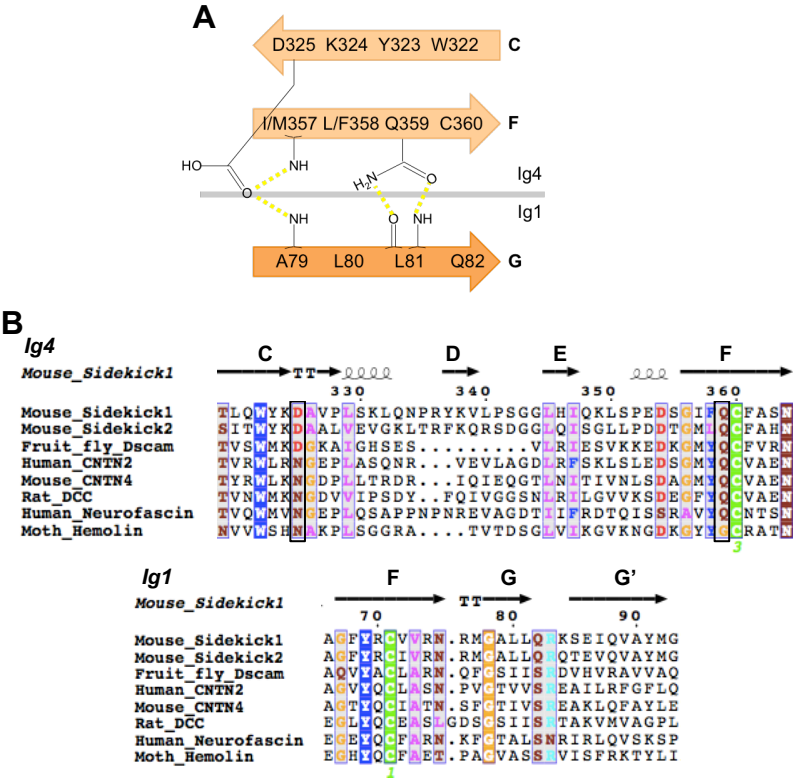


Figure 2–figure supplement 1. IgSF proteins containing horseshoe motifs including both Sdks have conserved intramolecular Ig1:Ig4 interactions

A. Schematic of the Sdk2 Ig1:Ig4 hydrogen-bonding network, which is conserved among horseshoe-containing IgSF proteins (Chen et al., 2013).

B. Multiple sequence alignment of horseshoe-containing IgSF proteins, for which there are crystal structures, showing the conserved Ig1 and Ig4 motifs from (A). Motif residues whose side chains form specific interactions are indicated by black boxes. The secondary structure elements from the Sdk structure are shown above the alignment.

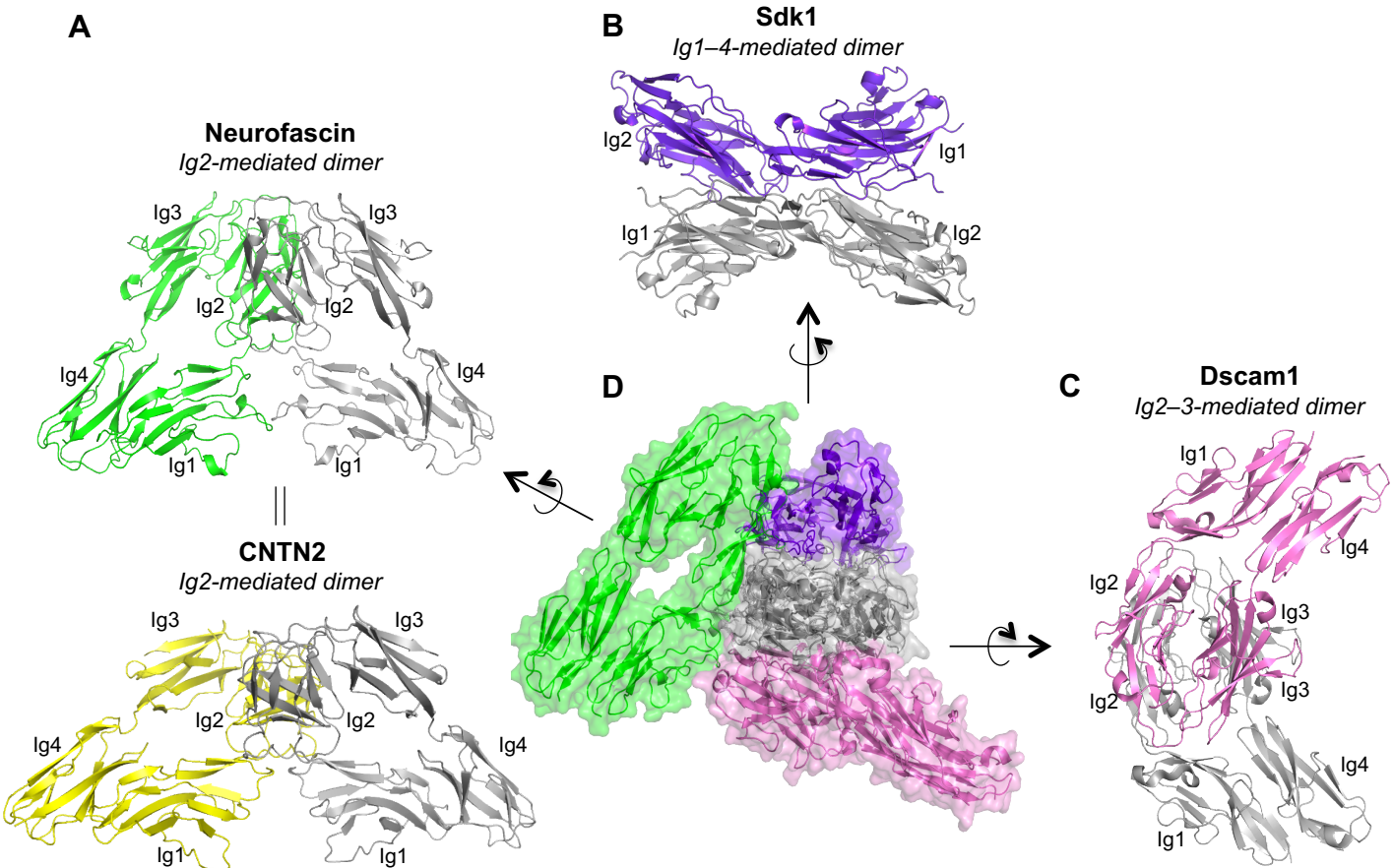


Figure 3—figure supplement 1. Dimer interactions mediated by IgSF horseshoe motif proteins

To date three different homophilic dimer conformations for IgSF proteins with Ig1–4 horseshoe motifs have been observed crystallographically and experimentally validated:

A. Neurofascin (PDB: 3P3Y) forms a symmetric Ig2-mediated dimer, primarily involving interactions between the Ig2 G-strands. The dimer structure is shown in ribbon depiction, with one protomer colored gray and the other green. The crystal structure of human CNTN2 (PDB: 2OM5, shown in the bottom panel with gray and yellow protomers) also shows a similar Ig2 G-strand mediated dimer arrangement, which was validated by mutagenesis experiments. The crystal structure of mouse CNTN4 (PDB: 3JXA) also shows a similar dimer arrangement in the crystal lattice.

B. As described in this paper Sdk1 and Sdk2 Ig1–4 regions form dimers mediated by all four Ig domains. The dimer observed in Sdk1_{Ig1–4} crystal form 2 (purple and gray protomers) is shown.

C. The horseshoe motif Ig1–4 regions of *Drosophila* Dscam1 (PDB: 3DMK, protomers colored gray and pink) form a symmetric dimer mediated by Ig2:Ig2 and Ig3:Ig3 interactions, however this dimer does not form in solution without the additional interactions mediated by Ig7 (Meijers et al., 2007).

D. Central panel shows a structural alignment of the gray protomer from each of the Neurofascin, Sdk1, and Dscam1 Ig1–4 dimer structures. Each of the proteins utilizes a different surface of the horseshoe motif to mediate homodimerization.

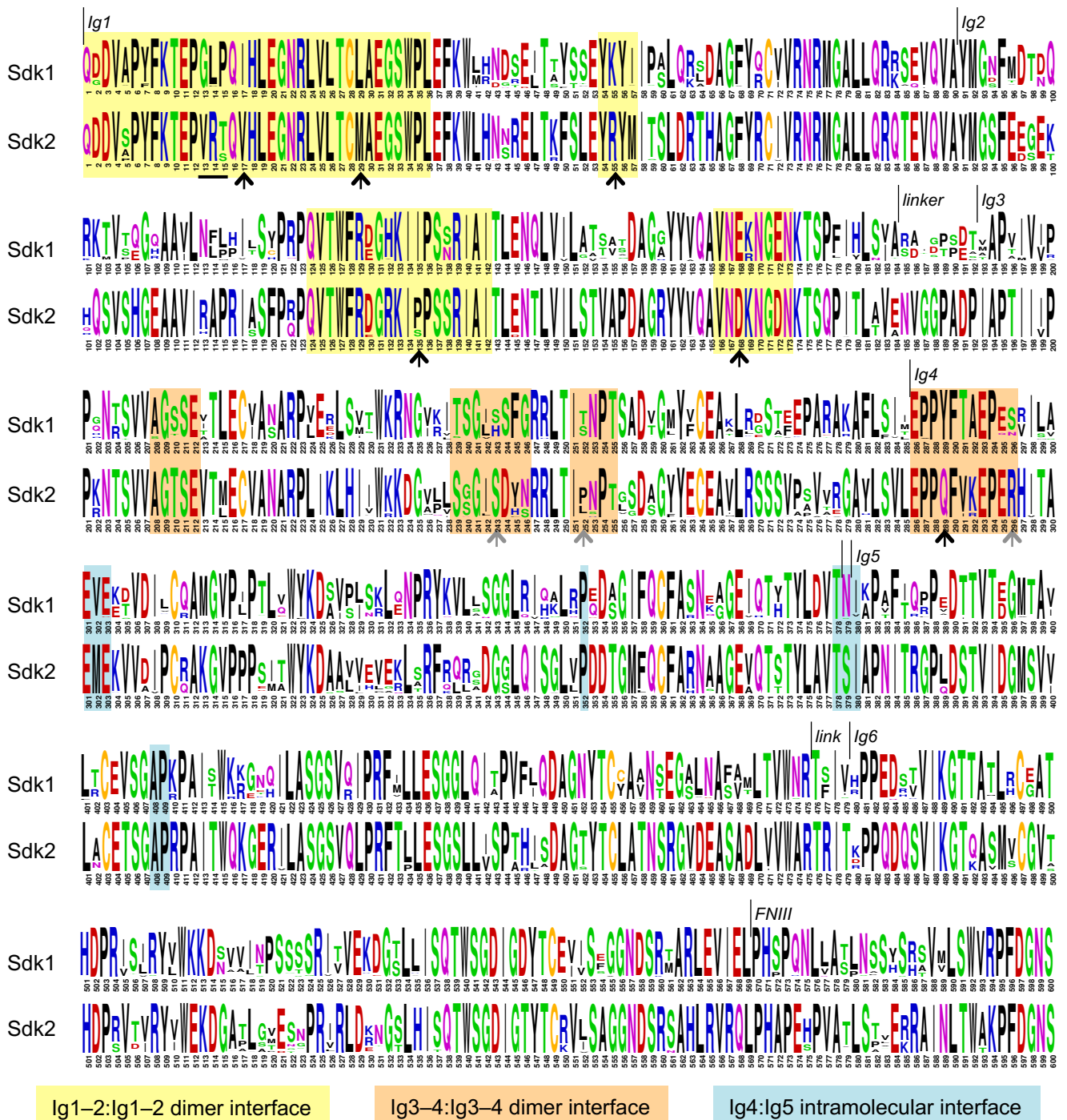


Figure 4—figure supplement 1. Sdk sequence logos showing conservation of the adhesive dimer interface residues and the Ig4:Ig5 intramolecular interface residues

Sequence logos for Sdk1 and Sdk2 generated from multiple sequence alignments of vertebrate Sdk1 and Sdk2 protein sequences (species used are listed in Table S5). Yellow highlights indicate Ig1-2:Ig1-2 dimer interface regions.

Orange highlights indicate Ig3-4:Ig3-4 dimer interface regions. Black arrows indicate key interface residues that show conserved differences between Sdk1 and Sdk2. Black underline indicates marginally interfacial residues with conserved differences. Grey arrows indicate key interface residues that differ between Sdk1 and Sdk2 but are not as well conserved. Cyan highlights indicate Ig4:Ig5 intramolecular junction interfacial residues.

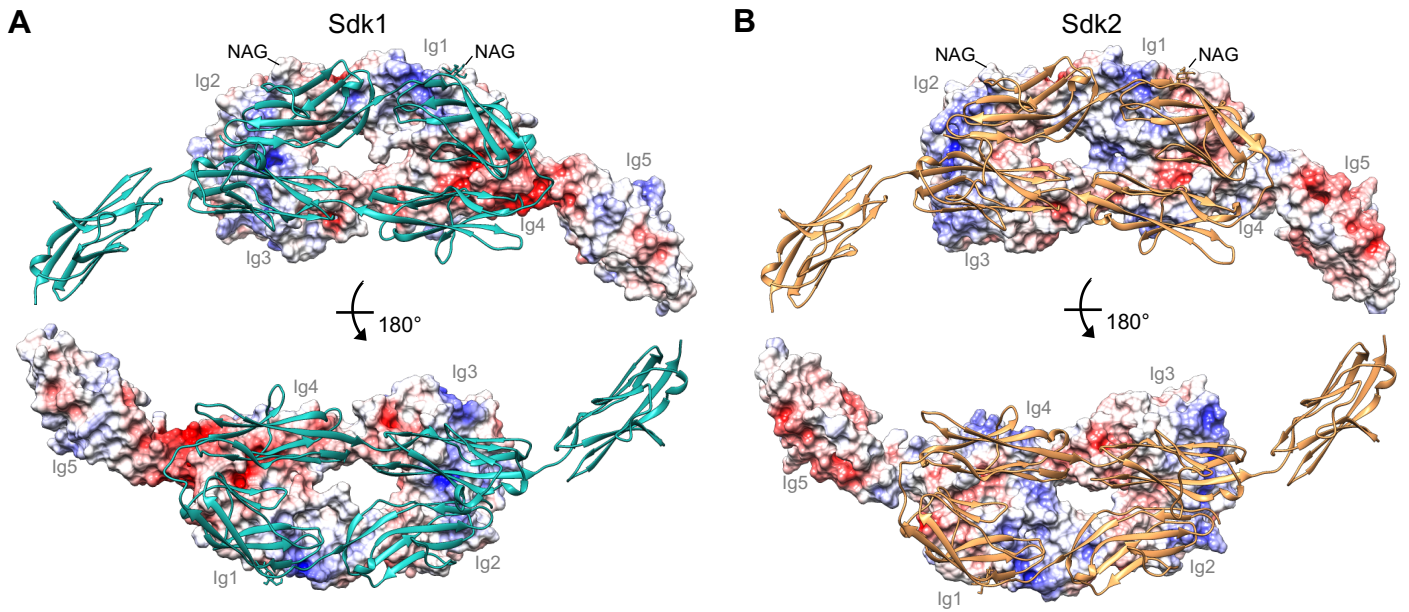


Figure 4—figure supplement 2. Electrostatic surfaces of the Sdk structures

A. Sdk1_{Ig1–5} dimer structure including the single glycan residue involved in the dimer interface (N-acetylglucosamine (NAG) bonded to asparagine 173). The top panel shows the electrostatic surface of the interacting face of chain A (Ig domains labeled), with chain B shown in ribbon depiction. The bottom panel shows the electrostatic surface of the interacting face of chain B (Ig domains labeled), with chain A shown in ribbon depiction. The Ig4–Ig5 linker region is highly negatively charged. Electrostatic potentials were calculated using DelPhi (Honig and Nicholls, 1995) and depicted at ± 5 keV using UCSF Chimera (Pettersen et al., 2004).

B. Sdk2_{Ig1–4} dimer structure (crystal form 1) including the single glycan residue involved in the dimer interface (NAG bonded to asparagine 173), with a model of Sdk2_{Ig5} based on Sdk1_{Ig1–5} produced using MODELLER (Webb and Sali, 2014). The top panel shows the electrostatic surface of the interacting face of chain A, with chain B shown in ribbon depiction. The bottom panel shows the electrostatic surface of the interacting face of chain B, with chain A shown in ribbon depiction.

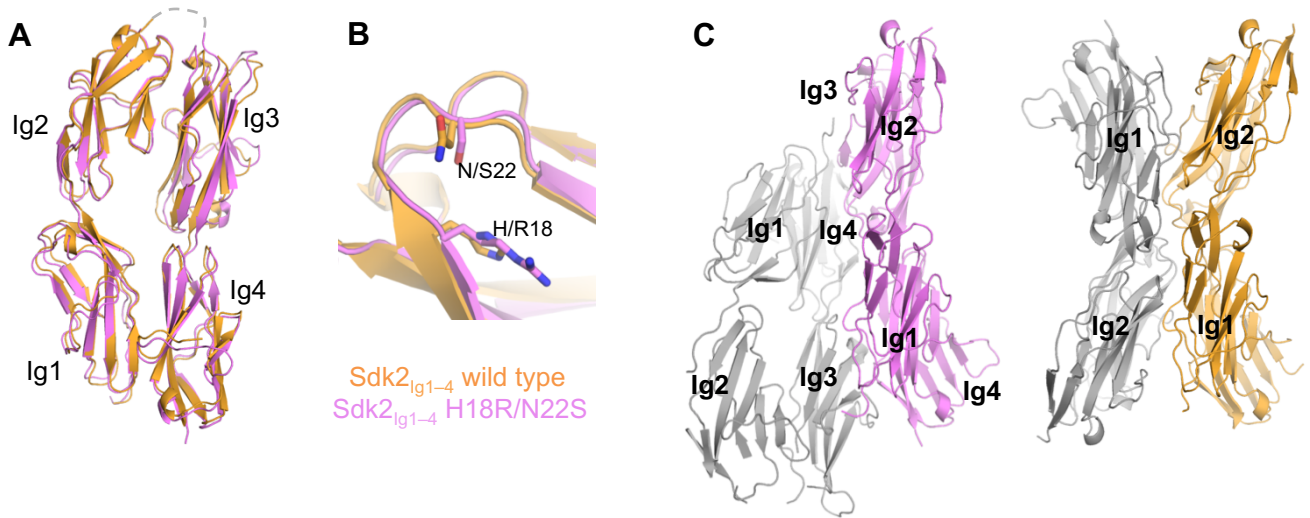


Figure 4—figure supplement 3. The crystal structure of Sdk2_{Ig1-4} H18R/N22S mutant shows complete loss of the wild-type dimer interaction

A. Structural alignment of individual protomers of the Sdk2_{Ig1-4} wild-type (crystal form 1, orange ribbon) and Sdk2_{Ig1-4} H18R/N22S mutant (pink ribbon) shows little difference between the overall horseshoe architecture (RMSD = 1.9Å).

B. Additionally the mutations do not cause a local conformational change in the AB loop.

C. However the wild-type dimer configuration with back-to-back horseshoes making extensive Ig1–2:Ig1–2 contacts is absent from the mutant crystal structure. The most extensive crystallographic dimer in the mutant crystal structure shows no Ig1–2:Ig1–2 contacts and shifted Ig3–4:Ig3–4 contacts relative to the wild-type dimer. Left: Sdk2_{Ig1-4} H18R/N22S crystallographic dimer with individual protomers in grey and pink; Right: Sdk2_{Ig1-4} wild-type dimer with individual protomers in grey and orange.

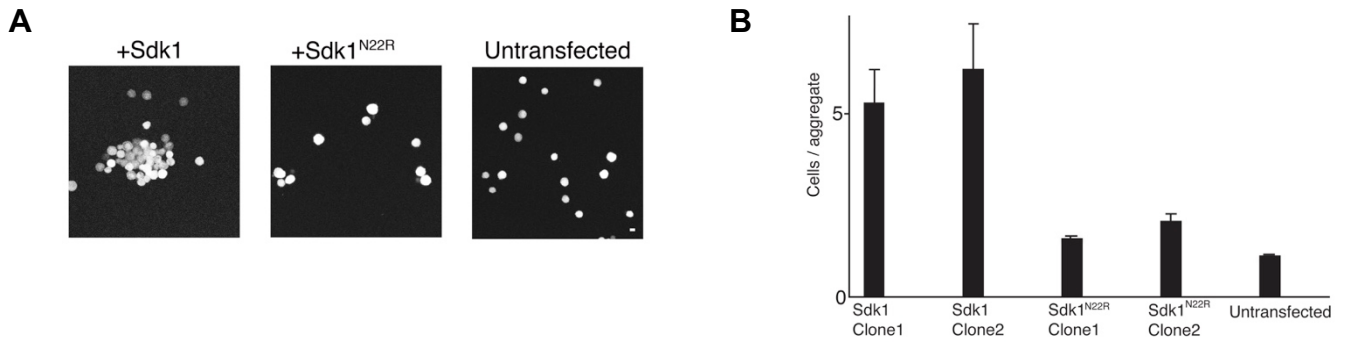


Figure 5—figure supplement 1. Cell aggregation assay showing the N22R dimer interface mutant impairs Sdk1-mediated cell adhesion

A. Aggregation assay of stably transfected L cell lines for wild-type Sdk1 and the Sdk1 N22R mutant.

B. Quantification of the aggregation assay for two stably transfected L cell lines for both wild-type Sdk1 and the Sdk1 N22R mutant (n=21–24, mean±S.E.).

Supplemental Tables

	Sdk1 _{Ig1-4} crystal form 1	Sdk1 _{Ig1-4} crystal form 2	Sdk1 _{Ig1-5}	Sdk2 _{Ig1-4} crystal form 1	Sdk2 _{Ig1-4} crystal form 2	Sdk2 _{Ig1-2} / Sdk1 _{Ig3-4}	Sdk2 _{Ig1-4} H18R/N22S
Data collection							
Date	7/12/2011	6/19/2010	4/5/2013	9/23/2013	10/8/2014	6/14/2014	7/19/2013
Beamline	NSLS X4A	NSLS X4C	NSLS X4C	NSLS X4C	APS 24-ID-C	APS 24-ID-E	APS 24-ID-E
Wavelength (Å)	1.743	1.005	0.979	0.979	0.979	0.979	0.979
Space group	C2	P3 ₂ 21	P2 ₁ 2 ₁ 2 ₁	P2 ₁	P2 ₁ 2 ₁ 2 ₁	P22 ₁ 2 ₁	I2
<i>Cell dimensions</i>							
a, b, c (Å)	163.07, 49.26, 60.73	158.19, 158.19, 53.53	70.14, 152.87, 158.97	55.11, 130.01, 78.39	82.64, 88.36, 106.55	78.35, 85.73, 146.35	110.89, 113.43, 185.91
α, β, γ (°)	90, 110.50, 90	90, 90, 120	90, 90, 90	90, 100.67, 90	90, 90, 90	90, 90, 90	90, 103.69, 90
Resolution (Å)	50.00–2.20 (2.24–2.20)	50.00–3.20 (3.29–3.20)	37.16–3.50 (3.69–3.50)	39.00–2.70 (2.85–2.70)	45.00–3.20 (3.37–3.20)	57.83–2.70 (2.85–2.70)	64.84–2.70 (2.85–2.70)
R _{merge}	0.100 (0.487)	0.147 (0.520)	0.132 (0.304)	0.099 (0.487)	0.122 (0.460)	0.144 (0.502)	0.101 (0.546)
I/σI	12.2 (2.4)	8.2 (1.5)	9.9 (3.9)	12.2 (2.8)	7.1 (2.4)	7.5 (3.1)	10.1 (2.5)
Completeness (%)	99.4 (100.0)	90.7 (72.9)	89.9 (79.5)	98.7 (98.5)	95.5 (96.5)	99.9 (100.0)	100.0 (100.0)
Redundancy	6.8 (6.6)	4.5 (3.6)	3.3 (2.7)	5.1 (5.1)	3.9 (3.8)	3.8 (3.9)	3.6 (3.6)
Refinement							
Resolution (Å)	38.19–2.20	34.74–3.20	36.99–3.50	39.00–2.70	44.78–3.20	45.37–2.70	39.52–2.70
Number of reflections	22884	11639	19853	29407	12640	27743	61442
R _{work} / R _{free} (%)	19.1 / 22.8	20.7 / 24.4	20.2 / 25.1	19.3 / 23.0	20.6 / 24.8	19.3 / 23.3	20.6 / 24.3
Molecules in ASU	1	1	2	2	2	2	4
<i>Number of residues</i>							
Protein	379	379	950	739	740	753	1478
Carbohydrate	4	6	13	7	8	4	12
Small molecule	0	0	2	1	0	3	0
Ion	7	6	3	12	0	2	14
Water	148	6	6	38	7	37	62
<i>R.m.s. deviations</i>							
Bond lengths (Å)	0.007	0.002	0.003	0.003	0.003	0.004	0.004
Bond angles (°)	0.880	0.489	0.613	0.705	0.713	0.715	1.064
<i>Ramachandran</i>							
Favored (%)	98.18	96.30	96.72	99.18	98.90	97.99	98.71
Allowed (%)	1.82	3.70	3.07	0.82	1.10	2.01	1.29
Outliers (%)	0.00	0.00	0.21	0.00	0.00	0.00	0.00
Wilson B	37.56	70.26	62.67	49.32	79.02	32.99	44.76
Overall B	45.03	75.68	64.80	62.20	77.57	33.95	60.67
PDB ID	5K6U	5K6V	5K6W	5K6X	5K6Y	5K6Z	5K70

Table S1. X-ray crystallography data collection and refinement statistics

Values in parentheses are for the outer shell. ASU = asymmetric unit; R.m.s. = Root mean square.

Protein	Chain	Ig14:Ig23 (°)	Ig1:Ig2 (°)	Ig2:Ig3 (°)	Ig3:Ig4 (°)	Ig1:Ig4 (°)	Ig4:Ig5 (°)
Sdk1 _{Ig1-4} crystal form 1	A *	30.0	40.6	23.2	27.8	27.8	
Sdk1 _{Ig1-4} crystal form 2	A *	37.1	43.5	23.1	33.1	35.6	
Sdk1 _{Ig1-5}	A *	31.7	42.8	23.4	28.3	30.4	54.3
	B	35.5	42.7	23.2	33.9	31.5	54.0
Sdk2 _{Ig1-4} crystal form 1	A	31.2	40.5	14.3	30.8	29.5	
	B *	30.9	42.3	18.2	30.8	27.4	
Sdk2 _{Ig1-4} crystal form 2	A *	26.2	42.5	18.8	26.7	27.4	
	B	26.0	41.5	17.4	35.7	27.9	
Sdk2 _{Ig1-2} / Sdk1 _{Ig3-4}	A	31.7	37.4	19.3	20.4	27.4	
	B *	22.1	41.5	22.7	26.7	26.9	
Sdk2 _{Ig1-4} H18R/N22S	A	24.7	45.3	23.3	19.5	32.7	
	B	19.3	44.2	24.9	17.8	29.2	
	C *	24.6	41.9	19.1	18.5	30.7	
	D	21.4	43.6	15.9	18.7	31.1	
Average		28.0 ± 5.4	42.2 ± 1.9	20.5 ± 3.3	26.3 ± 6.3	29.7 ± 2.5	54.2 ± 0.2

Table S2. Sidekick interdomain angles

Ig14:Ig23 is the angle between the Ig1:Ig4 and Ig2:Ig3 planes in the horseshoe. Ig1:Ig2 is the angle between the Ig1 and Ig2 axes, defined by mass weighting. *Indicates the chain shown in Figure 2A or S4A.

Protein	Ig14:Ig23 (°)	Ig1:Ig2 (°)	Ig2:Ig3 (°)	Ig3:Ig4 (°)	Ig1:Ig4 (°)	RMSD to Sdk1
Sdk1	34	42	23	31	31	N/A
Sdk2	29	42	17	31	28	1.0 Å
Dscam ^{PDB: 3DMK}	10	69	43	30	30	5.8 Å
CNTN2 ^{PDB: 2OM5}	8	43	21	13	29	3.4 Å
CNTN4 ^{PDB: 3JXA}	15	51	30	14	33	2.8 Å
DCC ^{PDB: 3LAF}	36	82	40	6	44	5.5 Å
Neurofascin ^{PDB: 3P3Y}	14	50	23	7	35	4.8 Å
Hemolin ^{PDB: 1BIH}	20	55	27	7	31	6.2 Å
Average	21 ± 11	54 ± 14	28 ± 9	17 ± 11	33 ± 5	

Table S3. Structural comparison of IgSF protein horseshoe structures

Average horseshoe protein interdomain angles over all chains in all crystal structures are given. Ig14:Ig23 is the average angle between the Ig14 and Ig23 planes. The right hand column gives the root mean square deviation over aligned Cα atoms (RMSD) between the Ig1-4 structures of the various IgSF proteins and Sdk1. The Ig1:Ig4 angle is most rigid among horseshoes, highlighting the importance of Ig1:Ig4 interface in horseshoe conformations.

Protein	Total interface BSA (Å ²)	Ig1-2:Ig1-2			Ig3-4:Ig3-4		
		BSA (Å ²)	H-bonds	Salt bridges	BSA (Å ²)	H-bonds	Salt bridges
Sdk1 _{Ig1-4} crystal form 1	1531	1531	2	2	0	0	0
Sdk1 _{Ig1-4} crystal form 2	3868	2693	16	2	915	8	0
Sdk1 _{Ig1-5}	3126	2917	18	2	209	0	0
Sdk2 _{Ig1-4} crystal form 1	3631	2337	16	2	1294	5	0
Sdk2 _{Ig1-4} crystal form 2	2899	2571	16	1	328	4	0
Sdk2 _{Ig1-2} / Sdk1 _{Ig3-4}	2525	2525	18	2	0	0	0

Table S4. Dimer interface characteristics

Interface statistics from the PISA server (Krissinel and Henrick, 2007).

Sidekick 1 sequences		Sidekick 2 sequences	
Accession	Species	Accession	Species
XP_686686.3	Danio rerio	XP_006029532.1	Alligator sinensis
XP_016122898.1	Sinocyclocheilus grahami	XP_004860637.1	Heterocephalus glaber
XP_006637301.2	Lepisosteus oculatus	XP_008165868.1	Chrysemys picta bellii
XP_012671777.1	Clupea harengus	XP_015090733.1	Vicugna pacos
XP_014801724.1	Calidris pugnax	XP_004749092.1	Mustela putorius furo
XP_010073520.1	Pterocles gutturalis	XP_011585064.1	Aquila chrysaetos canadensis
XP_010390686.1	Corvus cornix cornix	XP_009465259.1	Nipponia nippon
XP_011577492.1	Aquila chrysaetos canadensis	XP_007482760.1	Monodelphis domestica
XP_009318308.1	Pygoscelis adeliae	XP_015735584.1	Coturnix japonica
XP_010568603.1	Haliaeetus leucocephalus	XP_006970501.1	Peromyscus maniculatus bairdii
XP_008496763.1	Calypte anna	XP_013154663.1	Falco peregrinus
KFU86669.1	Chaetura pelagica	XP_005350737.1	Microtus ochrogaster
XP_014466304.1	Alligator mississippiensis	XP_003768600.1	Sarcophilus harrisii
KFQ99860.1	Nipponia nippon	NP_766388.2	Mus musculus
KFO91422.1	Buceros rhinoceros silvestris	XP_010719784.1	Meleagris gallopavo
XP_014436664.1	Pelodiscus sinensis	XP_011287640.1	Felis catus
XP_007065402.1	Chelonia mydas	XP_006247755.1	Rattus norvegicus
XP_015732346.1	Coturnix japonica	XP_013364113.1	Chinchilla lanigera
EOB00921.1	Anas platyrhynchos	XP_005597275.1	Equus caballus
KFP83217.1	Apaloderma vittatum	XP_014644694.1	Ceratotherium simum simum
XP_009090333.1	Serinus canaria	XP_011965263.1	Ovis aries musimon
XP_002190771.2	Taeniopygia guttata	XP_014695185.1	Equus asinus
XP_014122801.1	Zonotrichia albicollis	AAN15076.1	Gallus gallus
XP_005435700.1	Falco cherrug	XP_012517683.1	Propithecus coquereli
XP_005420504.2	Geospiza fortis	XP_010340778.1	Saimiri boliviensis boliviensis
XP_005145354.2	Melopsittacus undulatus	XP_006886456.1	Elephantulus edwardii
XP_009562886.1	Cuculus canorus	XP_015666575.1	Protophthalmos mucrosquamatus
XP_005054485.1	Ficedula albicollis	XP_004275482.1	Orcinus orca
XP_015498549.1	Parus major	XP_007185669.1	Balaenoptera acutorostrata scammoni
KFM12390.1	Aptenodytes forsteri	XP_511658.3	Pan troglodytes
XP_014725986.1	Sturnus vulgaris	XP_003813403.1	Pan paniscus
XP_007498480.1	Monodelphis domestica	XP_008010104.1	Chlorocebus sabaeus
XP_014108198.1	Pseudopodoces humilis	XP_007454451.1	Lipotes vexillifer
NP_989436.2	Gallus gallus	XP_011717998.1	Macaca nemestrina
XP_005229039.1	Falco peregrinus	XP_013002520.1	Cavia porcellus
KFR12275.1	Opisthocomus hoazin	XP_004041123.1	Gorilla gorilla gorilla
KFZ56915.1	Caprimulgus carolinensis	XP_014975718.1	Macaca mulatta
KFQ05823.1	Leptosomus discolor	XP_010719785.1	Meleagris gallopavo
XP_004386013.1	Trichechus manatus latirostris	NP_001138424.1	Homo sapiens
XP_006143656.1	Tupaia chinensis	XP_005584885.1	Macaca fascicularis
KFP21781.1	Egretta garzetta	XP_015135354.1	Gallus gallus
XP_005621203.1	Canis lupus familiaris	XP_015999084.1	Rousettus aegyptiacus
KFV61434.1	Picoides pubescens	XP_014133117.1	Falco cherrug
XP_004394197.2	Odobenus rosmarus divergens	XP_012418340.1	Odobenus rosmarus divergens
XP_005397914.2	Chinchilla lanigera	XP_004709379.1	Echinops telfairi
XP_008017012.1	Chlorocebus sabaeus	XP_005530779.2	Pseudopodoces humilis
XP_003469953.1	Cavia porcellus	XP_008010105.1	Chlorocebus sabaeus
XP_518946.3	Pan troglodytes	XP_016158177.1	Ficedula albicollis
NP_689957.3	Homo sapiens	XP_015501623.1	Parus major
XP_004045095.1	Gorilla gorilla gorilla	XP_016072154.1	Miniopterus natalensis
XP_007901677.1	Callorhynchus milii	XP_008154494.1	Eptesicus fuscus
XP_011742520.1	Macaca nemestrina	XP_014129483.1	Zonotrichia albicollis
XP_012494361.1	Propithecus coquereli	ELW68360.1	Tupaia chinensis
KFV75189.1	Struthio camelus australis	XP_012376244.1	Dasyatis novemcinctus
XP_007941083.1	Orycteropus afer afer	XP_003495730.1	Cricetulus griseus
XP_008017011.1	Chlorocebus sabaeus	KFO30522.1	Fukomys damarensis
XP_011812394.1	Colobus angolensis palliatus	XP_007641438.1	Cricetulus griseus
XP_011832325.1	Mandrillus leucophaeus	XP_008054805.1	Tarsius syrichta
EHH17116.1	Macaca mulatta	XP_006754687.1	Myotis davidii
EHH51984.1	Macaca fascicularis	XP_015445099.1	Pteropus alecto
XP_003795492.1	Otolemur garnettii	XP_012588441.1	Condylura cristata
NP_808547.3	Mus musculus		
XP_004671382.2	Jaculus jaculus		
XP_008696690.1	Ursus maritimus		
XP_006859935.1	Chrysocloris asiatica		
XP_006889806.1	Elephantulus edwardii		
XP_014962337.1	Ovis aries musimon		
XP_007174719.1	Balaenoptera acutorostrata scammoni		

Table S5. Protein amino acid sequences used to generate Sdk sequence logos