Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Structural Basis for the self-recognition of sDSCAM in Chelicerata

Jie Cheng
Sichuan University
Yamei Yu
Sichuan University
Xingyu Wang
Sichuan University
Ting Liu
Sichuan University
Daojun Hu
Sichuan University
Yongfeng Jin
Zhejiang University https://orcid.org/0000-0001-7125-7055
Ying Lai
Sichuan University
Tian-Min Fu
The Ohio State University https://orcid.org/0000-0002-6281-1752
Qiang Chen (qiang_chen@scu.edu.cn)
Sichuan University https://orcid.org/0000-0002-2398-4482

Article

Keywords:

Posted Date: August 29th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1973770/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Structural Basis for the self-recognition of sDSCAM in Chelicerata

Abstract

To create a functional neural circuit, neurons develop a molecular identity to discriminate self from non-self. The invertebrate Dscams and vertebrate Pcdhs are implicated in determining synaptic specificity. Recently identified in Chelicerata, a shortened Dscam (sDscam) has been shown to resemble the isoform-generating characters of both Dscam and Pcdh and represent an evolutionary transition. Here we presented the molecular details of sDscam self-recognition via both *trans* and *cis* interactions using X-ray crystallographic data, functional assays and mathematical analysis. Based on our results, we proposed a molecular zipper model for the assemblies of sDscam to mediate cell-cell recognition. In this model, sDscam utilized FNIII domain to form side-by-side interactions with neighboring molecules in the same cell while established hand-in-hand interactions via Ig1 domain with molecules from another cell around. Together, our study provided a framework for understanding the assembly, recognition, and evolution of sDscam.

Introduction

The complexity of eukaryotic nervous system is established via neuronal cell interactions with vast diversity and specificity. During development, neurons need to discriminate self from non-self to establish appropriate connections. The neuronal wiring and self-avoidance rely on the extraordinary recognition diversity of cell surface molecules, such as the *Drosophila* Down syndrome cell adhesion molecules (Dscams)[1,2] and the mammalian clustered protocadherins (Pcdhs)[3]. *Drosophila* Dscam1 encodes 38,016 distinct isoforms via mutually exclusive RNA splicing[1], while human clustered Pcdh generates 52 isoforms using alternative promoters[4]. Recently, a shortened *Dscam* (*sDscam*) gene family with tandemly arrayed 5′ cassettes in Chelicerata species has been identified, which encodes ~50 to 100 isoforms[5,6]. With high sequence similarities to *Drosophila* Dscam1 and a striking gene organizational resemblance to the 5' variable region of vertebrate clustered Pcdhs, Chelicerata sDscam expanded its isoform diversity via a combination of alternative promoter and RNA splicing selections, representing remarkable functional convergence of invertebrate Dscams and vertebrate Pcdhs[5].

Both invertebrate Dscam and vertebrate Pcdh isoforms exhibit isoform-specific *trans* homophilic interactions to generate avoidance signals[2,3]. *Drosophila* Dscam1 uses three variable Ig domains (Ig2, Ig3, and Ig7) to perform homophilic binding[7] while mammalian Pcdh's homophilic recognition is mediated by N-terminal EC1–EC4 Domains[8]. Chelicerata sDscams could be classified into two subfamilies, sDscamα and sDscamβ, containing one and two variable Ig domains, respectively[5]. Different from the *Drosophila* Dscam1 that contains 10 immunoglobulin (Ig) domains and 6 fibronectin type III (FNIII) repeats, Chelicerata shortened sDscam only has 3 Ig domains and 3 FNIII domains in its extracellular part, corresponding to Ig7-Ig9 and FNIII1, 2, 5 of Dscam1, respectively[5] (Fig. 1). How this shortened sDscam fulfill its functional roles remains elusive.

Given the huge discrepancy in isoform diversity between Dscam (38,016 isoforms in

Drosophila) and Pcdh (52 isoforms in human), there is a big evolutionary gap between insects and vertebrates. Chelicerata sDscam, which resembles the isoform-generating features of both invertebrate Dscam and vertebrate Pcdh, provides us with a unique opportunity to get a glimpse of the evolutionary transitions.

Here, we present 12 crystal structures of sDscam fragments Ig1, Ig1-2, Ig1-3, FNIII1, FNIII2, FNIII3, FNIII2-3, and FNIII1-3, establishing the binding behavior of the *trans* (cell-to-cell) and *cis* (the same cell surface) binding modes for Chelicerata sDscams. Crystal structures and cell aggregation assays demonstrated that both α and β subfamilies of sDscam used Ig1 for the isoform-specific *trans* recognition in a hand-in-hand manner. Furthermore, the *cis* interactions were established by the three FNIII domains of sDscam, which assemble into a cross fold with a kink between FNIII2 and FNIII3. Combined with mathematical analysis, we proposed a zipper-like model for the assemblies formed by the entire sDscam ectodomain between cells. This study provided insights into sDscam-mediated cell discrimination and deepened our understandings of the evolutionary scenario of cell recognition molecule diversity.

Results

Structure determination of sDscam

A shortened sDscam in Chelicerata combines alternative promoter and RNA splicing to generate diverse isoforms, resembling both insect Dscams and vertebrate Pcdhs[5] (Fig. 1). In chelicerate *Mesobuthus martensii*, *sDscam* gene family contains tandemly arrayed cassettes, which comprise two or four exons encoding a single or two variable Ig domains, respectively (Fig. S1), and thus can be classified as sDscam α and β subfamilies. In the current study, we determined the crystal structures of sDscam from both α and β subfamilies to provide molecular details for its *trans* and *cis* recognitions.

A recent study showed that the *trans* homophilic interactions of Chelicerata sDscam were mediated via its N-terminal Ig1 domain[9]. To elucidate the molecular details of *trans* interactions of sDscam, we determined the crystal structures of Ig1 from *Mesobuthus martensii* sDscam isoforms α1, α7, and β6v2. To validate Ig2 and Ig3 are not involved in the *trans* homodimerization of sDscam, we further determined the crystal structures of the fragments Ig1-2 (isoform β2v6) and Ig1-3 (isoform α25) from *M. martensii*. We also determined the crystal structures of sDscam Ig1 from another chelicerate species *Limulus polyphemus* (isoforms α7 and β3v7) to provide insights into the conservation of sDscam *trans* recognition within Chelicerata. These crystal structures of Ig fragments were determined at 1.3 ~ 3.1 Å resolutions (Table S1 & S2).

It has been speculated that Chelicerata sDscam could form *cis* multimers via membrane proximal FNIII domains [9]. To analyze the *cis* interactions of sDscam, we determined the crystal structures of different FNIII fragments of *M. martensii* sDscam: FNIII1, FNIII2 and FNIII3 domains of isoform α7, and FNIII2-3 and FNIII1-3 fragments of isoform β2v6. These crystal structures of FNIII fragments were determined at 1.4 ~ 3.2 Å resolutions (Table S1 & S2).

Structural basis for the *trans* homodimerization of sDscam

The crystal structures of sDscam Ig1 from both *M. martensii* and *L. polyphemus* revealed a common assembly of antiparallel homodimer for both α and β subfamilies of sDscam (Fig. 2a). In addition, all these Ig1 homodimers resemble each other well with root mean square deviation (RMSD) values of 1.0 to 1.6 Å, indicating a shared mechanism of sDscam *trans* recognition within Chelicerata (Fig. 2b). As the homologue of *Drosophila* Dscam1 Ig7, sDscam Ig1 adopted a very similar conformation as *Drosophila* Dscam1 Ig7 homodimer (Fig. 2c), highlighting a conserved mechanism for the *trans* recognition in Insecta Dscam and Chelicerata sDscam. The fragments Ig1-2 (isoform β2v6) and Ig1-3 (isoform α25) both formed *trans* homophilic dimers exclusively through interactions mediated by the Ig1 domain (Fig. 2d). The Ig2 and Ig3 domains in Ig1-2 and Ig1-3 linearly followed after Ig1 domain without participating *trans* interactions.

The Ig fold is featured by a sandwich of two β-sheets, denoted as ABED face and CFG face, respectively. A previous study has demonstrated that *Drosophila* Dscam1 Ig7 use its ABED face to form a homophilic dimer through a charge complementary mechanism to ensure an antiparallel binding mode[7]. Similarly, sDscam Ig1 also employs the ABED face to form an antiparallel homodimer (Fig. 2). However, among the seven Ig1 structures reported here, the complementary electrostatic potential surface pattern (negative in one end, neutral in the middle, and positive in the other end) has been only observed in the three β isoforms, but not in any one of the four α isoforms (Fig. 3a & S2). In addition, the electrostatic potential arrangement of sDscam Ig1 ABED face was opposite to that of *Drosophila* Dscam1 Ig7 (Fig. 3b).

More interestingly, in some Ig1 crystal structures (*M. martensii* α1, α7, and *L. polyphemus* α7), the two protomers within a homodimer were related via a 2-fold symmetry axis, namely, these two protomers had identical conformations. This implied that the homophilic *trans* interactions of Ig1 were also symmetry-related.

To reveal the principles of Ig1 interactions, we analyzed all the interfaces of Ig1 homodimers. Structural analysis showed that B and E strands constituted the core of ABED face and dominated the *trans* homophilic interactions of sDscam Ig1. The residue preceding the conserved B-strand cystine always interacts with its counterpart of the other protomer. Electrostatic and shape complementarity had been observed in all Ig1 dimer interfaces (Fig. 3c). The electrostatic repulsion and/or shape non-complementarity may avoid the binding between different sDscam isoforms, resulting in isoform-specific interactions. Consistently, the sequence alignment of 40 isoforms of *M. martensii* sDscamα subfamily showed low conservation for the ABED face, but high conservation of CFG face (Fig. 3d). In particular, the β strands of an Ig domain have alternate distribution of inner-facing and outer-facing residues, and we observed a much higher variability of the outer-facing residues that mediate Ig1 dimerization, especially Ig1 E strand (Fig. 3d). Moreover, all the Ig1 isoforms without exception have a potential N-linked glycosylation site (NXS/T) located at the F strand (Fig. 3d). This glycosylation prevents the CFG face from

establishing homophilic interactions and helps define the interface of the sDscam homodimer.

Isoform-specific inter-cell recognition of sDscam

To validate the functional significance of the *trans* recognitions of sDscam observed in the crystal structures, we measured the ratio of aggregated cells mediated by wild type and truncated sDscams overexpressed in cells using a cell aggregation assay (Fig. 4a). For both α and β subfamilies, wild type sDscams lead to cell aggregation, while truncation of Ig1 domain, but not other Ig domains, completely abolished cell aggregation (Fig. 4b). This finding supported the functional significance of Ig1 in mediating the *trans* recognition of sDscam, which was consistent with our structural analysis (Fig. 2).

To validate the isoform specific interaction mediated by sDscam, we examined cell aggregation using cells expressing different isoforms of sDscam. For each isoform of *M. martensii* sDscam, we fused mCherry and EGFP to its C-terminus respectively and evaluated the binding specificity by pairwise isoform combinations (Fig. 4c). Cells expressing the same isoforms of sDscam formed aggregation while cells with different isoforms of sDscam were separated from each other (Fig. 4d & 4e). These data showed that all the tested isoforms exhibited a strict *trans* homophilic specificity.

Detection of the *cis* polymerization of sDscam on living cell surface

A recent study suggested that Chelicerata sDscam utilized its membrane proximal FNIII domains to form *cis* interactions[9]. To identify the specific FNIII domain responsible for *cis* interactions, we developed a living cell-based assay. In this assay, the intracellular portion of the receptor tyrosine kinase cKIT and a Flag tag were fused to the FNIII domains of sDscam to make a sDscam-cKIT chimera (Fig. 5a). As monomeric cKIT cannot phosphorylate itself, the neighboring cKIT molecules phosphorylate each other when they assemble into oligomers. The phosphorylation intensity is correlated to quantities of oligomerized molecules. By monitoring the cKIT tyrosine phosphorylation through anti-phospho-tyrosine immunoblotting in SF9 cells, we can quantify the number of oligomerized sDscam molecules.

We found that FNIII2 alone but not FNIII1 or FNIII3 could promote cKIT tyrosine phosphorylation (Fig. 5b), suggesting the essential role of FNIII2 domain in mediating sDscam *cis* interactions. Furthermore, significantly stronger cKIT tyrosine phosphorylation signal was observed in FNIII1-3 fragment, compared to FNIII2 alone, FNIII12, or FNIII2-3 fragments (Fig. 5c), indicating that the three FNIII domains cooperate to establish an optimal *cis* recognition of sDscam.

Structural basis for the *cis* homodimerization of sDscam

To elucidate the structural basis for the *cis* interactions of sDscam, we determined the crystal structures of different FNIII fragments of *M. martensii* sDscam: FNIII1, FNIII2 and FNIII3 domains of isoform α7, and FNIII2-3 and FNIII1-3 fragments of isoform β2v6. In contrast to the elongated conformation of sDscam Ig1-3, FNIII1-3 adopted an “L” shape

resulting from a kink between the FNIII2 and FNIII3 domains (Fig. 6a). Similarly, the kink between FNIII2 and FNIII3 was also observed in the crystal structure of FNIII2-3 fragment (Fig. 6a). Notably, there were seven FNIII2-3 molecules per asymmetric unit, which all adopted a very similar kink conformation with root-mean-square deviation (RMSD) of 0.727 Å over 198 Ca atoms (Fig. S3). Thus, this kink conformation may reflect a physiological geometric assembly. The inter-domain interface of the kink contains a salt bridge formed by R404 and D578, and the hydrophobic interaction mediated by L492 and L496 (Fig. 6b).

In contrast to the previously proposed parallel model[9], two L-shaped sDscam FNIII1-3 molecules formed a homodimer with a cross configuration (Fig. 6b). The dimer interface buries 2,109 Å^2 of surface area, which is composed of one FNIII2-FNIII2 interface, two FNIII1-FNIII2 interfaces and two FNIII1-FNIII3 interfaces (Fig. 6b). Consistent with our cellular data (Fig.5), FNIII2 played a pivotal role in sDscam *cis* recognition as it formed intermolecular interactions with FNIII1 and FNIII2 of the other protomer and intramolecular interactions with FNIII3 of the same protomer (Fig. 6c).

The FNIII2-FNIII2 interface was located at the center of the *cis* homodimer, where P395, P417 and Y420 formed a hydrophobic core (Fig. 6c). Y420 appeared to be essential because it forms a hydrogen bond with P395 and π–π interactions with the symmetry-related Y420. The FNIII1-FNIII2 interface was dominated by hydrogen bond networks formed by the side-chain of N357 and Q485, and the main-chain of L358 and E486 (Fig. 6c). In contrast, the FNIII1-FNIII3 interface was composed of both hydrophobic and hydrophilic interactions (Fig. 6c).

Oligomerization status of sDscam Ig and FNIII domains in solution

To biochemically characterize the oligomerization status of sDscam Ig and FNIII domains *in vitro*, we compared the gel filtration profiles of different sDscam Ig or FNIII fragments (Fig. S4). In the gel-filtration profile, the Ig1 peak was eluted earlier than the Ig2 or Ig3 peak, indicating that the Ig1 domain assembles into larger oligomers than Ig2 or Ig3 in solution. The FNIII domain has a similar molecular size and shape as those of the Ig domain. The single FNIII domains of sDscam were eluted at a similar position with that of Ig2 or Ig3, while FNIII2-3 was eluted at a position corresponding to the Ig1 homodimer. These data suggested that each single FNIII domain and the FNIII2-3 fragment existed as monomers in solution. The FNIII1-3 was eluted at a very similar position as that of Ig1-3, indicating the FNIII1-3 also existed as a homodimer in solution. The gel-filtration analysis supported our observations in the crystal structures: Ig1 mediated the formation of *trans* homodimer (Fig. 2) and all the three FNIII domains were involved in the *cis* interactions (Fig. 6).

Theoretical Analysis of sDscam-mediated neuronal diversity

To avoid incorrect recognition between non-self neurites, some fraction of common isoforms between different neurons must be tolerated without triggering recognition[10]. A zipper model has been proposed for Pcdh intermembrane assembly[8], which provides a chain-termination mechanism to achieve an extremely high common-isoform tolerance.

This explains how the much fewer Pcdh isoforms may be sufficient to mediate neuronal discrimination in vertebrates.

Here we used Monte-Carlo simulations[10] to calculate the probabilities that a pair of neurons randomly express the same isoform(s), assuming 10 Pcdh or sDscam isoforms expressed per cell (Fig. S5). To obtain a low enough probability of incorrect non-self recognition (for example, $1x10^{-8}$), the common-isoform tolerance would be 80~90%. With the given total number of possible isoforms and common-isoform tolerance (insect Dscam1: 5,000, 20%; scorpion sDscam: 95, 80%; mouse Pcdh: 58, 80%), we also calculated the probabilities that two neurons randomly expressing different numbers of isoforms will incorrectly recognize each other as self (Fig. 7a). The calculation results demonstrated the high common-isoform tolerance as a key factor to mediate non-self discrimination with a limited number of isoforms. It is worth noting that for a small isoform repertoire (such as mouse Pcdh, 58 isoforms), the probability of incorrect recognition increased as too many isoforms expressed per cell (Fig. 7a), indicating a smaller isoform diversity may favor less isoforms expressed per cell. This is in accordance with the reported number of isoforms expressed by single neuron: some tens of isoforms for Dscam1[11,12] vs 10~15 isoforms for Pcdh[13]. Considering the comparable isoform number of clustered Pcdh and sDscam, it is plausible to speculate that Chelicerata sDscam adopts the zipper-like assembly as clustered Pcdh.

Model of the full extracellular domain assembly of sDscam

Our crystal structures of sDscam Ig and FNIII fragments cover the entire sDscam ectodomain and enable us to propose a model for the panorama of sDscam *trans* and *cis* recognitions. A distance of about 24 nm of the *Drosophila* Dscam1-mediated cell adhesion interface has been observed by electron microscopy[14]. We supposed that sDscam-mediated cell interface had a similar distance as that of *Drosophila* Dscam1, and the calculation arrived at 53° for the angle between sDscam Ig and FNIII domains (Fig. 7b).

Based on the crystal structures of *M. martensii* sDscam Ig1-3 *trans* dimer (isoform α25) and FNIII1-3 *cis* dimer (isoform β2v6), we proposed a molecular zipper model for sDscam recognition (Fig. 7b & 7c). On cell surface, two neighbouring sDscam molecules form a cross *cis*-dimer via the FNIII domains in a side-by-side manner, while each extended arm (Ig domains) of a sDscam *cis*-dimer engages in *trans* with two different *cis*-dimers on an apposed cell in a hand-in-hand manner. Thus, a continuous array assembled through alternating *cis* and *trans* interactions of sDscam. In order to incorporate into the zipper-like assembly, the incoming sDscam *cis*-dimer must have one isoform to match the isoform on the exposed end of the zipper assembly (Fig. 7c). The zipper assembly will terminate when a matched isoform is not available, resembling the chain-termination model for Pcdh-mediated discrimination between self and non-self[8,15].

Discussion

In the nervous system, individualization and specialization of the cells are at the basis of their function, and thus a vast diversity of cell surface recognition proteins is a necessity. Alternative splicing of pre-messenger RNA is the most common mechanism to achieve the

diversification[16]. To create diversified repertoires of recognition, different RNA splicing mechanisms have been evolved, such as mutually exclusive RNA splicing for fruit fly Dscam1[1] and alternative promoters for human Pcdh[4]. Chelicerata sDscam has been found to combine alternative promoters and RNA splicing to generate its isoform diversity [5], and therefore offers a unique perspective on the diversity and evolution of the neural cell surface receptors.

Our structural analysis combined with biochemical and cellular assays provide a molecular zipper model for Chelicerate sDscam. In this model, there are two types of interactions: the *trans* interactions and the *cis* interactions. The *trans* interactions of Chelicerata sDscam were mediated exclusively through the Ig1 domain for both sDscamα and sDscamβ subfamilies (Fig. 2 & 4). It is worth noting that sDscamβ subfamily has two variable Ig domains, Ig1 and Ig2. Although not involved in the *trans* interaction, the variable Ig2 domain may play other as-yet unknown functions, such as immune regulation observed for Arthropod Dscams[17,18]. The *cis* interactions are mediated by the FNIII domains that are constant within the sDscam subfamilies. We evaluated the *cis* polymerization of sDscam FNIII domains on living cell surface (Fig. 5) and determined the crystal structure of sDscam homo *cis*-dimer (Fig. 6). However, whether hetero *cis* interactions exit among sDscam subfamilies has not been evaluated due to the limitation of these assays. It has been reported that different sDscam subfamilies may form *cis*-dimers in co-IP assays[9]. Different isoforms within sDscamβ subfamily share a 41 to 61% identity of protein sequence for the FNIII part, while sDscamα and sDscamβ FNIII domains share a 35 to 42% protein sequence identity. Sequence alignment of these FNIII domains showed conservation of the interface residues, to a certain extent, among different subfamilies (Fig. S6). Whether hetero *cis*-dimers exit in sDscam and if so, whether they adopt the same conformation as the homo *cis*-dimer need further studies.

Neuronal self-avoidance requires a reliable discrimination between self and non-self neurites. Two encountering neurites from different cells should not incorrectly recognize each other as self. This would be simple when the two cells express completely different subsets of sDscam isoforms, since no homophilic binding would occur. However, there is a counterintuitively high probability of overlap of randomly expressed isoforms in different neurons, even with a very large repertoire of the isoform pool. For example, using Monte-Carlo simulations, it has been shown that, with 50 isoforms randomly selected per neuron from a pool of 20,000, the probability that two neurons express at least one common isoform is about 12%[10]. Therefore, some fraction of common isoforms between different neurons must be tolerated without triggering recognition.

It has been demonstrated that thousands of isoforms are necessary for Dscam1, with an approximately 20% tolerance, to provide enough number of neurons with functionally unique identities[10]. Compared to the thousands of isoforms of Dscam1, clustered Pcdh only has tens of isoforms and thus have to exploit a different recognition mechanism from Dscam1. A zipper model has been proposed for Pcdh intermembrane assembly[8], which has been observed by X-ray crystallography and cryo-electron tomography[19]. This zipper

model provides a chain-termination mechanism: even a single mismatch is sufficient to disrupt the chain extension and thus to significantly reduce the number of favorable interactions. Therefore, an extremely high common-isoform tolerance is achieved. It has been shown that even four of five expressed Pcdh isoforms being identical could avoid incorrect non-self recognition[20], implying a common-isoform tolerance of at least 80%, which is in line with our mathematical calculation (Fig. S5).

The isoform diversity varies enormously in insect Dscam and vertebrate Pcdh, and therefore two different strategies have been used for cell recognition. *Drosophila* Dscam1 with a vast isoform repertoire appears to perform *trans* interactions as monomers, while mammalian Pcdh with a limited isoform diversity forms *cis* multimers to further assemble a zipper-like structure between cell membranes[8,19,20]. The cryo-electron tomography revealed that Pcdh forms zipper-like assemblies in a native-mimicking membrane environment[19]. However, such continuous ordered assemblies have not been observed for *Drosophila* Dscam1 on cell surface[14]. Although Chelicerata sDscam shared a high sequence homology to invertebrate Dscam and used Ig1 domain to mediate *trans* interaction mimicking Dscam1 Ig7 domain (Fig. 2), its limited isoform diversity implied that it likely used the zipper-like assembly strategy for cell recognition as vertebrate clustered Pcdh. Our mathematical calculation (Fig. S5 & 7a) and the *trans* and *cis* interactions observed in the crystal structures of sDscam enabled us to propose a molecular zipper model for sDscam self-recognition (Fig. 7c). A switch-like repulsion response might be elicited when the extension of the zipper reaches a threshold.

Arthropoda, the largest phylum in the animal kingdom, accounts for more than 80% of known animal species. Chelicerata is the second largest subphylum of arthropods, the evolutionary history of which could extend back to the Cambrian (~ 524 million years ago)[21,22]. The origin of s*Dscam* genes has been proposed to lie before the split of Arachnida and Merostomata, and phylogenetic analysis suggests that the sDscam may originate from the sequential shortening of the canonical Dscam[5,6]. Dscam1 has been implicated as a Netrin-1 receptor and the Netrin-1 binding site is located at the Ig7-Ig9 fragment[23,24] Interestingly, the three Ig domains of Chelicerata sDscam correspond to Dscam1 Ig7-Ig9[5], implying that the shortened sDscam keeps the Netrin-1 binding site in evolution and plays a similar role as Dscam1 in neural development.

Diverged from other Arthropod lineages ~500 million years ago, Chelicerata represented the insect Dscam homologue, sDscam, which has a genomic organization similar to vertebrate clustered Pcdh. Our structural and functional results reported here showed that sDscam Ig1 mediated *trans* interaction resembling insect Dscam1 Ig7, while the membrane-proximal FNIII domain mediated *cis* interaction to form a zipper-like assembly mimicking the vertebrate clustered Pcdh. Together, these studies advanced our understanding of sDscam and shed light on the evolutionary landscape of the recognition molecule diversity.

Materials and Methods

Protein production and crystallization

The sDscam Ig1-2, Ig1-3 constructs and *M. martensii* isoform α7 were expressed in a baculovirus system (BacMagic, Novagen), and other Ig fragments were expressed in *E. coli* strain SHuffle (NEB). All the FNIII fragments were expressed in *E. coli* strain Rosetta (DE3) (Novagen). The protein purification was performed as previously described[25]. Briefly, the target proteins were purified by IDA-Nickel magnetic beads (BeaverBeads™, Beaverbio), followed by tag-removing using TEV protease, and gel filtration (Superdex 10/300 GL, GE Healthcare). Crystals were grown at 16 °C using the hanging-drop vapor diffusion method. The crystallization conditions were summarized in Table S1.

Data collection and structure determination

Diffraction data were collected on beamline BL19U1 at the Shanghai Synchrotron Radiation Facility (SSRF). Collected data were processed by HKL3000[26]. The crystal structures of single Ig or FNIII domains were determined by molecular replacement using the predicted structures from AlphaFold[27] as search models. The multidomain structures were determined by molecular replacement using the structures of single domains as search models. Structure refinement and model building were performed with PHENIX[28] and Coot[29]. All models were validated with MolProbity[30]. Details of the data processing and refinement statistics were summarized in Table S2. All structure figures were prepared with ChimeraX [31] and PyMOL (www.pymol.org).

Cell aggregation assay

Cell aggregation assay was performed as previously reported with minor modifications [9]. Sf9 cells were cultured in SIM SF medium (Sino Biological) supplemented with 10% fetal bovine serum (Gibco) at 27 °C. Baculoviruses were obtained according to the manufacturer's instructions of BacMagic transfection kit (Novagen). Sf9 cells were seeded at $8x10^5$ cells per well in a 6-well plate and infected with P2 recombinant viruses followed by 2-day incubation at 27 °C. Then cells were collected by centrifuge, washed once by 1 × HCMF (Coolaber) gently and resuspended with 1 × HCMF. For cell aggregation assay, the 6-well plate preincubated by 1% BSA in 1 × HBSS (Solarbio) at 4 °C overnight and washed with 1 × HCMF. Then 2 × 10^5 cells in 2 ml 1 × HCMF were transferred into each well of the 6-well plate. Cell suspension in 6-well plates was incubated at 27 °C in gyratory shaker at 60 rpm for 1 h. Samples were then observed with inverted fluorescence microscope.

Tyrosine phosphorylation assay

Sf9 cells were cultured in SIM SF medium (Sino Biological) supplemented with 10% fetal bovine serum (Gibco) at 27 °C. For infection, cells were seeded at $2x10^6$ cells per well in a 6-well plate, infected with P2 recombinant viruses followed by 3-day incubation at 27 °C. Cells were then collected by centrifuge, washed twice with PBS, solubilized in lysis buffer containing 120 mM NaCl, 25 mM HEPES (pH 7.4), 1 mM EGTA, 0.75 mM $MgCl_2$, 10% glycerine, 1% triton, 1 mM NaF, 2 mM sodium orthovanadate, protease inhibitor cocktail (EDTA-free) and incubated on ice for 1 h. The lysates were centrifuged at 12,000 rpm for 20 min at 4 °C and then the supernatants were incubated with anti-Flag magnetic beads (Smart Lifesciences) for 2 h at 4°C with rotating. The beads were washed twice with

washing buffer containing 50 mM HEPES (pH 7.4), 150 mM NaCl, 1 mM EGTA, 0.75 mM $MgCl_2$, 10% glycerol, 0.1% triton, and 2 mM sodium orthovanadate. Finally, the samples were boiled for 5 min and analyzed by Western Blot with anti-Flag (Smart Lifesciences) and anti-phosphotyrosine (HuaBio) antibodies. Protein was visualized by chemiluminescence imager ChemiDoc Touch (Bio-Rad) and quantification were performed using ImageJ program.

Statistical analysis

Statistical analysis was performed using the GraphPad Prism 8 Software. Two-tailed Student's t-test was used to compare differences. Significance level was set at P>0.05. All values were reported as means±s.d.

Stochastic modelling

Given different number of isoforms (*m*) stochastically expressed from total isoform pool available (*i*), the probability (*Q*) that a pair of neurons has more than *t* isoforms in common is calculated using the following formula [20]:

$$Q_{i,m,t} = 1 - \sum_{k=0}^{t} \binom{m}{k} \frac{\binom{i-m}{m-k}}{\binom{i}{m}}$$

For Figure S5, we assumed 10 distinct isoforms (*m*) are stochastically expressed per cell from the total isoform pool available, the probability of a pair of neurons sharing *t* or more isoforms was calculated.

For Figure 7a, the tolerance (*t*) was assumed to be 20% for Dscam1 and 80% for Pcdh and sDscam. Then for a given number of isoforms expressed per cell (*m*), *t* was calculated as the integer part of 0.2 × m or 0.8 × m.

Declaration of interests

The authors declare no competing interests.

Reference

1. Schmucker, D. et al. Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* **101**, 671-84 (2000).
2. Wojtowicz, W.M. et al. A vast repertoire of Dscam binding specificities arises from modular interactions of variable Ig domains. *Cell* **130**, 1134-45 (2007).
3. Lefebvre, J.L., Kostadinov, D., Chen, W.V., Maniatis, T. & Sanes, J.R. Protocadherins mediate dendritic self-avoidance in the mammalian nervous system. *Nature* **488**, 517-21 (2012).
4. Wu, Q. & Maniatis, T. A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell* **97**, 779-90 (1999).
5. Yue, Y. et al. A large family of Dscam genes with tandemly arrayed 5' cassettes in Chelicerata. *Nat Commun* **7**, 11252 (2016).
6. Cao, G. et al. A chelicerate-specific burst of nonclassical Dscam diversity. *BMC Genomics* **19**, 66 (2018).
7. Li, S.A., Cheng, L., Yu, Y., Wang, J.H. & Chen, Q. Structural basis of Dscam1

homodimerization: Insights into context constraint for protein recognition. *Sci Adv* **2**, e1501118 (2016).
8. Rubinstein, R. et al. Molecular logic of neuronal self-recognition through protocadherin domain interactions. *Cell* **163**, 629-42 (2015).
9. Zhou, F. et al. Chelicerata sDscam isoforms combine homophilic specificities to define unique cell recognition. *Proc Natl Acad Sci U S A* **117**, 24813-24824 (2020).
10. Hattori, D. et al. Robust discrimination between self and non-self neurites requires thousands of Dscam1 isoforms. *Nature* **461**, 644-8 (2009).
11. Zhan, X.L. et al. Analysis of Dscam diversity in regulating axon guidance in Drosophila mushroom bodies. *Neuron* **43**, 673-86 (2004).
12. Neves, G., Zucker, J., Daly, M. & Chess, A. Stochastic yet biased expression of multiple Dscam splice variants by individual cells. *Nat Genet* **36**, 240-6 (2004).
13. Mountoufaris, G., Canzio, D., Nwakeze, C.L., Chen, W.V. & Maniatis, T. Writing, Reading, and Translating the Clustered Protocadherin Cell Surface Recognition Code for Neural Circuit Assembly. *Annu Rev Cell Dev Biol* **34**, 471-493 (2018).
14. Guo, L. et al. Structure of cell-cell adhesion mediated by the Down syndrome cell adhesion molecule. *Proc Natl Acad Sci U S A* **118**(2021).
15. Rubinstein, R., Goodman, K.M., Maniatis, T., Shapiro, L. & Honig, B. Structural origins of clustered protocadherin-mediated neuronal barcoding. *Semin Cell Dev Biol* **69**, 140-150 (2017).
16. Maniatis, T. & Tasic, B. Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* **418**, 236-43 (2002).
17. Ng, T.H. & Kurtz, J. Dscam in immunity: A question of diversity in insects and crustaceans. *Dev Comp Immunol* **105**, 103539 (2020).
18. Ng, T.H., Chiang, Y.A., Yeh, Y.C. & Wang, H.C. Review of Dscam-mediated immunity in shrimp and other arthropods. *Dev Comp Immunol* **46**, 129-38 (2014).
19. Brasch, J. et al. Visualization of clustered protocadherin neuronal self-recognition complexes. *Nature* **569**, 280-283 (2019).
20. Thu, C.A. et al. Single-cell identity generated by combinatorial homophilic interactions between alpha, beta, and gamma protocadherins. *Cell* **158**, 1045-1059 (2014).
21. Dunlop, J.A. Geological history and phylogeny of Chelicerata. *Arthropod Struct Dev* **39**, 124-42 (2010).
22. Lozano-Fernandez, J. et al. A molecular palaeobiological exploration of arthropod terrestrialization. *Philos Trans R Soc Lond B Biol Sci* **371**(2016).
23. Liu, G. et al. DSCAM functions as a netrin receptor in commissural axon pathfinding. *Proc Natl Acad Sci U S A* **106**, 2951-6 (2009).
24. Ly, A. et al. DSCAM is a netrin receptor that collaborates with DCC in mediating turning responses to netrin-1. *Cell* **133**, 1241-54 (2008).
25. Cheng, L. et al. DCLK1 autoinhibition and activation in tumorigenesis. *Innovation (N Y)* **3**, 100191 (2022).
26. Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes. *Acta Crystallogr D Biol Crystallogr* **62**, 859-66 (2006).
27. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**,

583-589 (2021).
28. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21 (2010).
29. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32 (2004).
30. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010).
31. Goddard, T.D. et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **27**, 14-25 (2018).
32. Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G. & Gibson, T.J. Multiple sequence alignment with Clustal X. *Trends Biochem Sci* **23**, 403-5 (1998).
33. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320-4 (2014).

Figures and figure legends

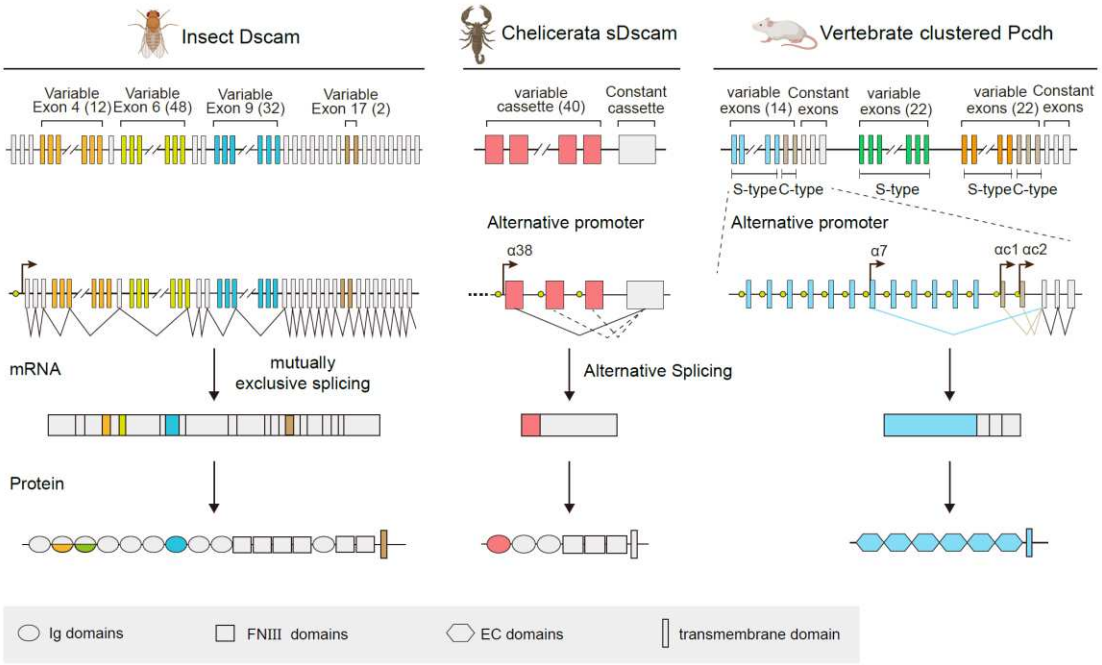


Figure 1. Schematic depiction of Dscam, sDscam and Pcdh gene organization and isoform diversity. *Drosophila melanogaster* Dscam, *Mesobuthus martensii* sDscam and mouse Pcdh have been chosen as samples. The variable regions are indicated by different colors, while the common regions are colored in grey. The promoters are represented as yellow filled circles. Dscam applies mutually exclusive RNA splicing to generate tens of thousands of isoforms. For Pcdh, each variable exon is preceded by a promoter and only the cap-proximal exon engages in the splicing with the first constant exon. The shorten sDscam uses both alternative promoter and RNA splicing to generate its isoform diversity. For clarity, only sDscamα subfamily is shown. More details of sDscam isoforms are shown in Fig. S1.

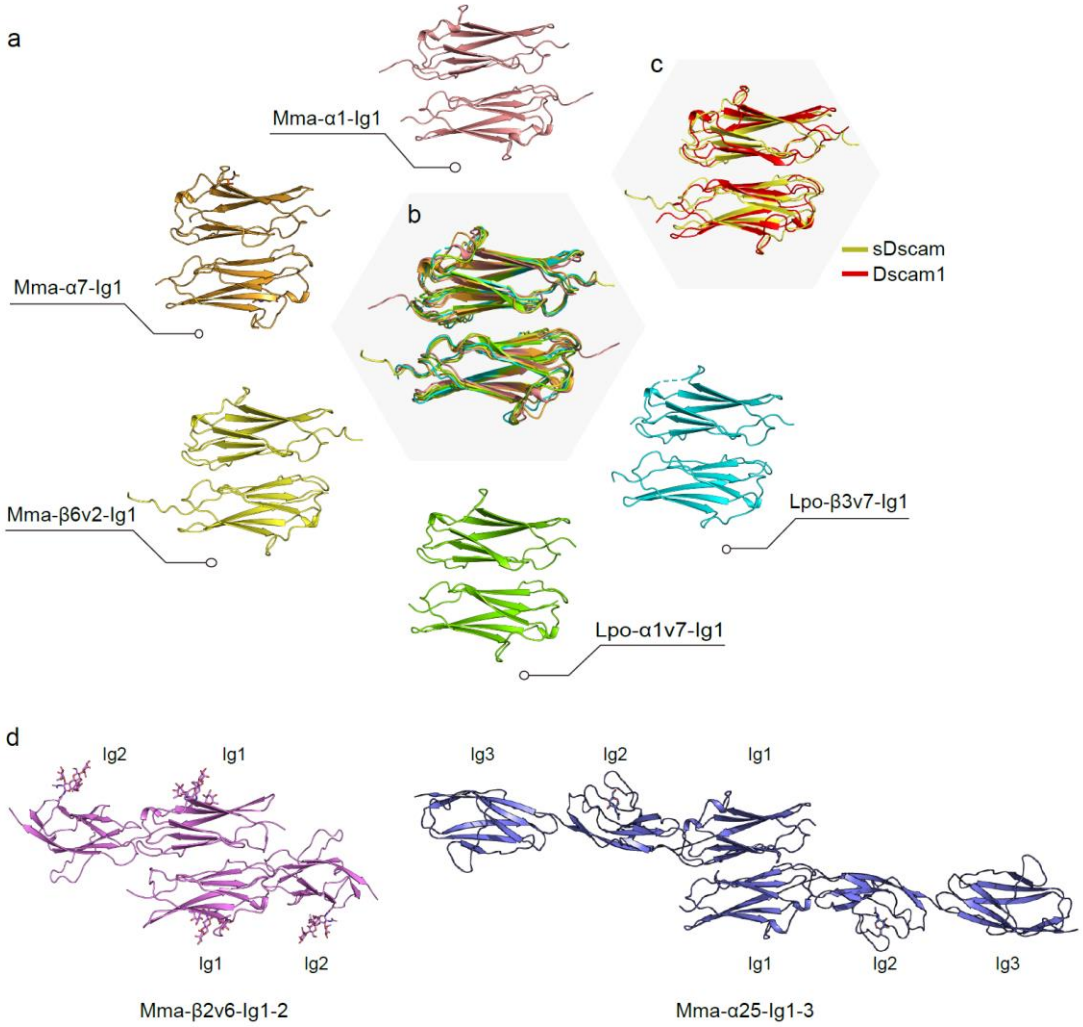


Figure 2. The homodimer of sDscam Ig1. (**a**) The crystal structures of Ig1 homodimers from *M. martensii* and *L. polyphemus*. Glycans are shown in stick representation. (**b**) Superposition of sDscam Ig1 homodimers. (**c**) Superposition of the homodimers of sDscam Ig1 and Dscam1 Ig7. (**d**) The crystal structures of *M. martensii* Ig1-2 and Ig1-3 homodimers.

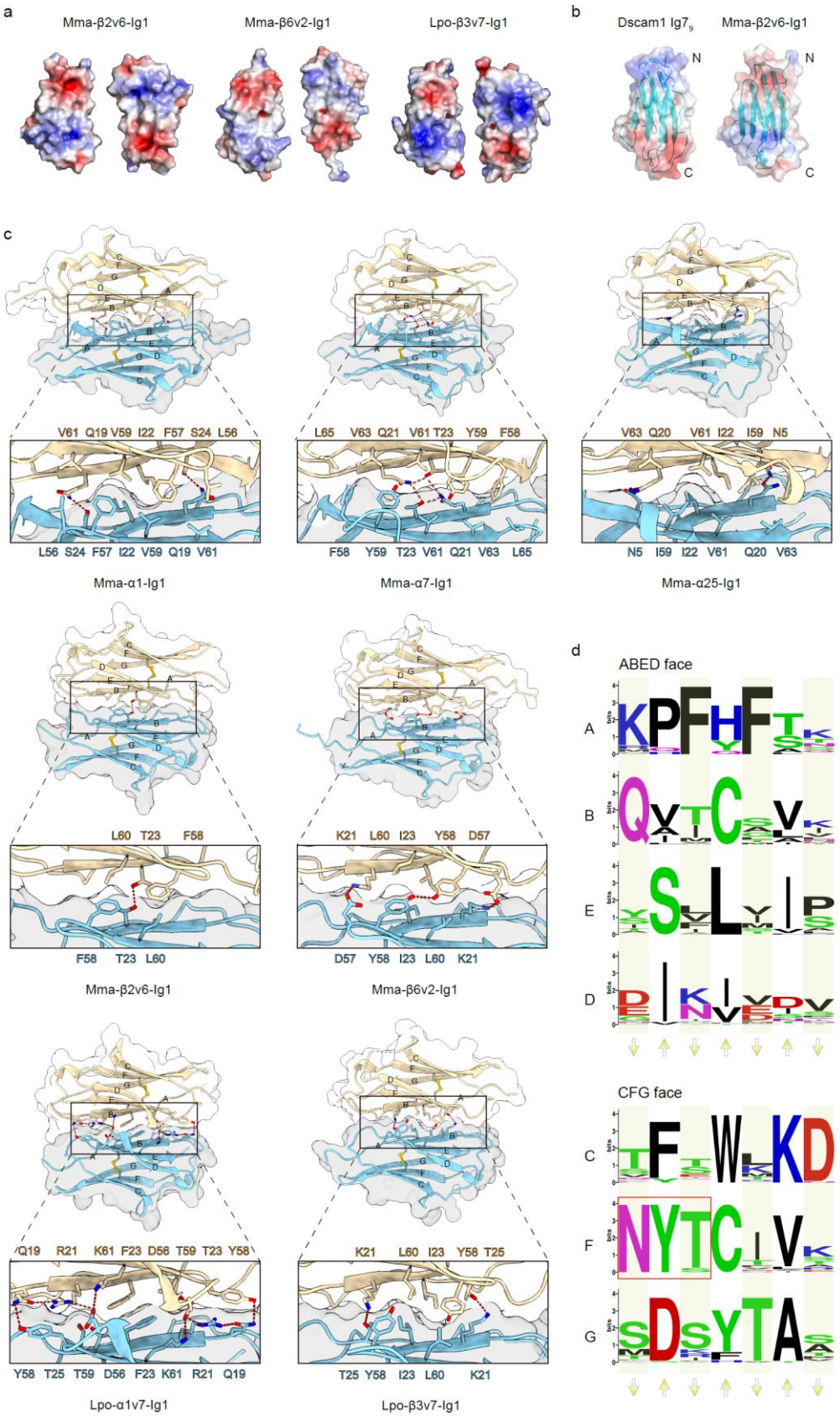


Figure 3. The interface of Ig1 homodimer. (**a**) Open-book view of the electrostatic surface potential of the homodimers of Ig1 isoforms. Blue and red (±5 kT/e) indicate the positively and negatively charged areas of the protein, respectively. (**b**) The electrostatic

surface potential patterns of Dscam1 Ig7 and sDscam Ig1 display opposite orientations. The N-terminus and C-terminus of the Ig domain are indicated. (**c**) The Ig1 homodimer interfaces are viewed down the two-fold symmetry axis. The two protomers within an Ig1 homodimer are colored in wheat and cyan, respectively. The details of the interface are shown in the enlarged part. (**d**) Sequence logo representation of the conservation of Ig1 isoforms from *M. martensii* sDscam. A total of 40 isoforms of the sDscam α subfamily are used for comparison. The up arrows and down arrows indicate the inner-facing and outer-facing residues, respectively. The outer-facing residues are shaded. The constant N-glycosylation site (NXT/S) is indicated by a red box.

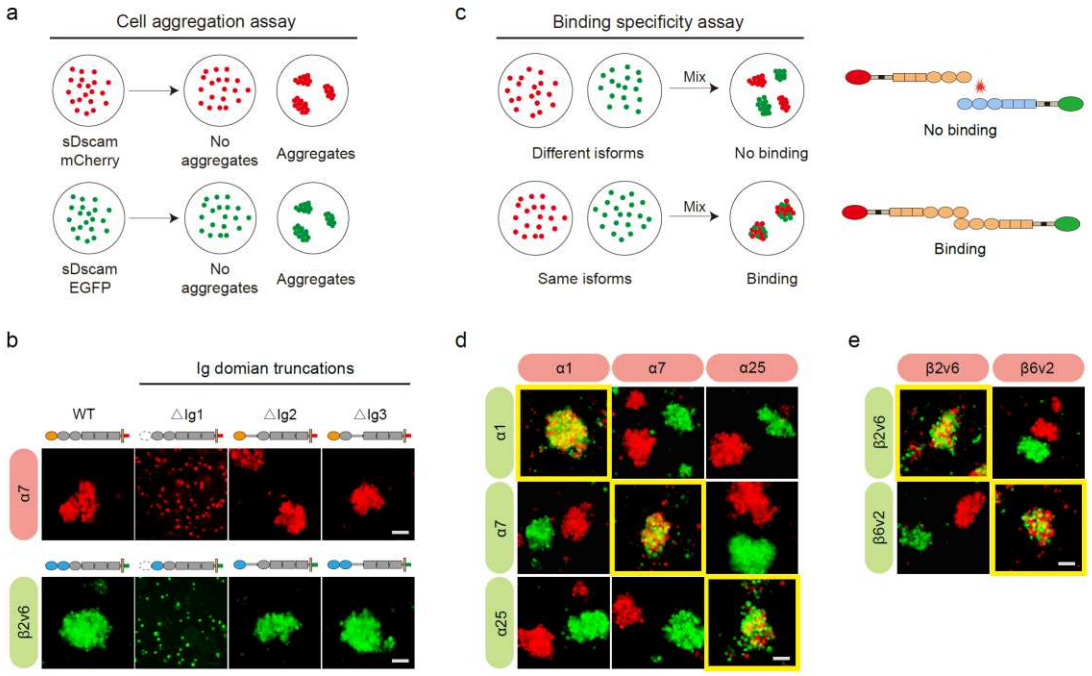


Figure 4. The Ig1 mediates the isoform-specific homophilic interactions of sDscam. (**a**) Schematic diagram of the cell aggregation assay. The mCherry- or EGFP-tagged sDscam proteins are expressed on Sf9 cells to their ability to induce cell aggregation. (**b**) The wild type or truncation sDscam proteins are expressed on Sf9 cells to assay for their ability to form cell aggregates. (**c**) Schematic diagram of the binding specificity assay. Cells expressing mCherry- or EGFP-tagged sDscam isoforms are mixed and assayed for homophilic or heterophilic binding. The *trans* interaction between cells expressing sDscam isoforms with different tags is indicated by mixed red-green coaggregation, while the negative results are indicated by red-green segregation. (**d**) Pairwise combinations within three sDscamα isoforms are assayed for their binding specificity. (**e**) Pairwise combinations within two sDscamβ isoforms are assayed for their binding specificity. Scale bar, 100 μm.

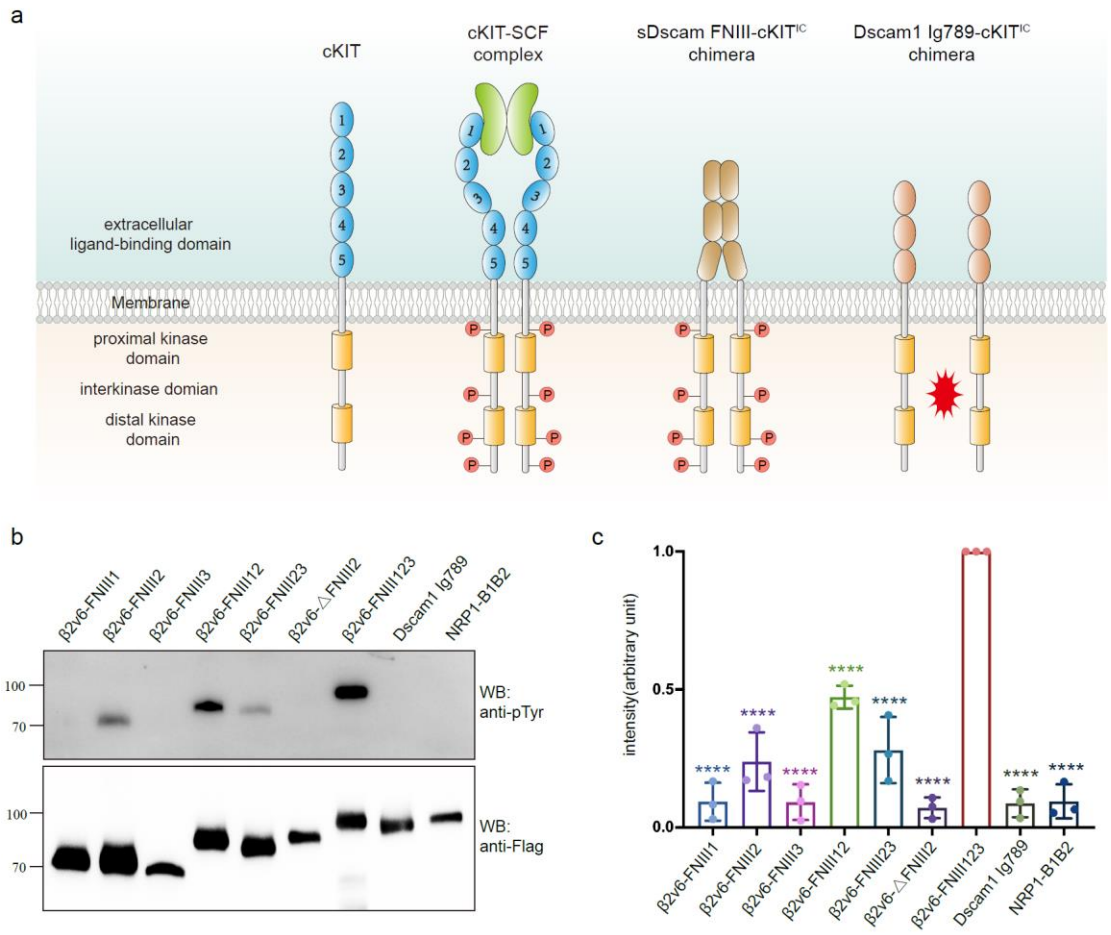


Figure 5. Detection of sDscam *cis* interactions on living cell surface. (**a**) Schematic diagram of the phosphorylation dimerization assay. The sDscam-cKIT chimera are expressed on Sf9 cell surface. If dimerization occurs, the intracellular portion of cKIT gets phosphorylated, and the phosphorylation signals are then monitored through anti-phospho-tyrosine immunoblotting. (**b**) Immunoblotting analysis of cKIT tyrosine phosphorylation. The same samples were detected through anti-phospho-tyrosine and anti-Flag immunoblotting to monitor the phosphorylation protein and the total protein, respectively. Upper panel, anti-phospho-tyrosine immunoblotting. Lower panel, anti-Flag immunoblotting. The Ig7-9 fragment of *Drosophila* Dscam1 and a nonrelative protein (human NRP1 B1B2 domain) are used as negative controls. For ΔFNIII2 construct, the intermediate FNIII2 domain is replaced by a 15-residue GGS linker. (**c**) Quantification of the cKIT tyrosine phosphorylation. The intensities of the bands in (**b**) are calculated. The phosphorylation signals (anti-phospho-tyrosine immunoblotting intensities) are normalized by the protein level of each sample (anti-Flag immunoblotting intensities). The value of FNIII1-3 is set as 100%. Values are means + SD (n = 3). ****P<0.0001.

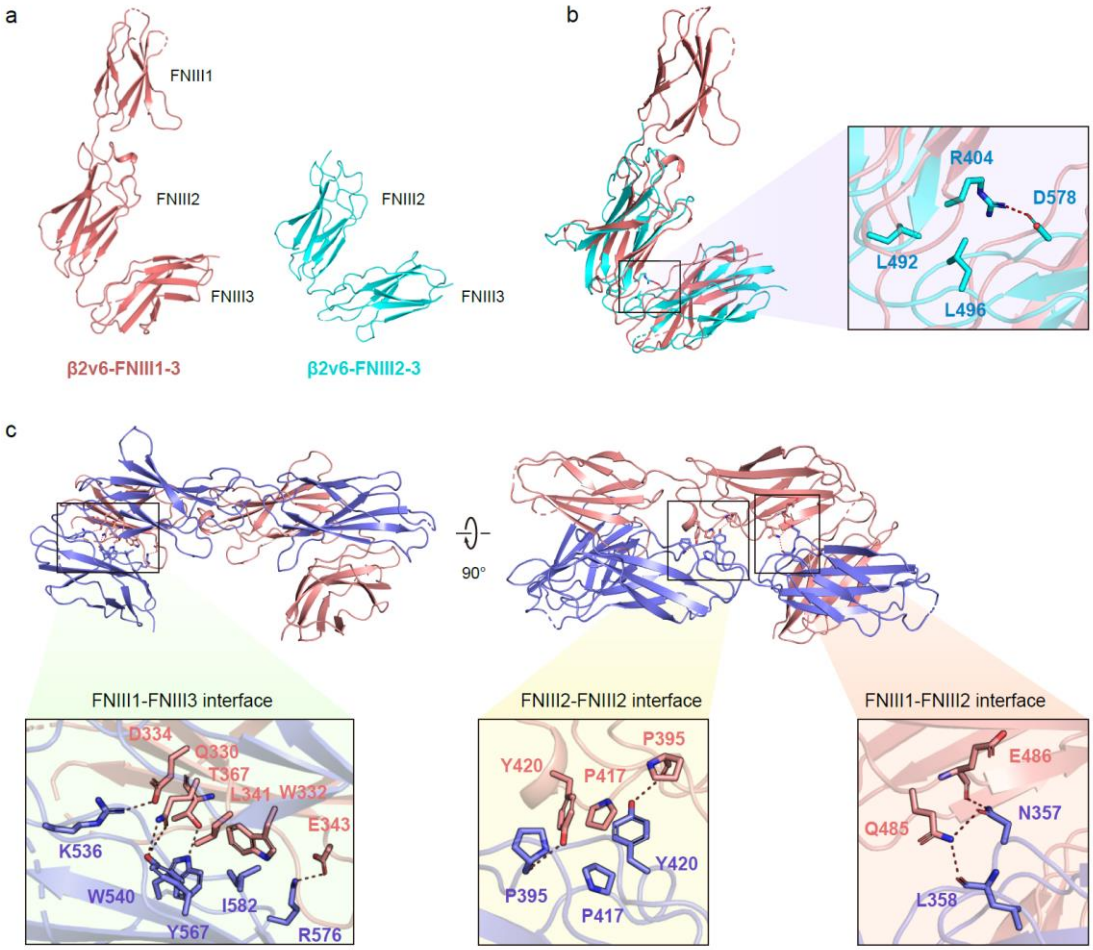


Figure 6. The homodimer of sDscam FNIII domains. (**a**) The three FNIII domains of sDscam adopt an L-shaped configuration. A kink between FNIII2 and FNIII3 are observed in both FNIII1-3 (red) and FNIII2-3 (cyan) structures. (**b**) Superposition of the crystal structures of sDscam FNIII1-3 (red) and FNIII2-3 (cyan). The residues involved in the inter-domain interactions between FNIII2 and FNIII3 are shown as stick and labeled. (**c**) The homodimer of sDscam FNIII domains adopt a cross configuration. The details of three interfaces are shown in the enlarged part.

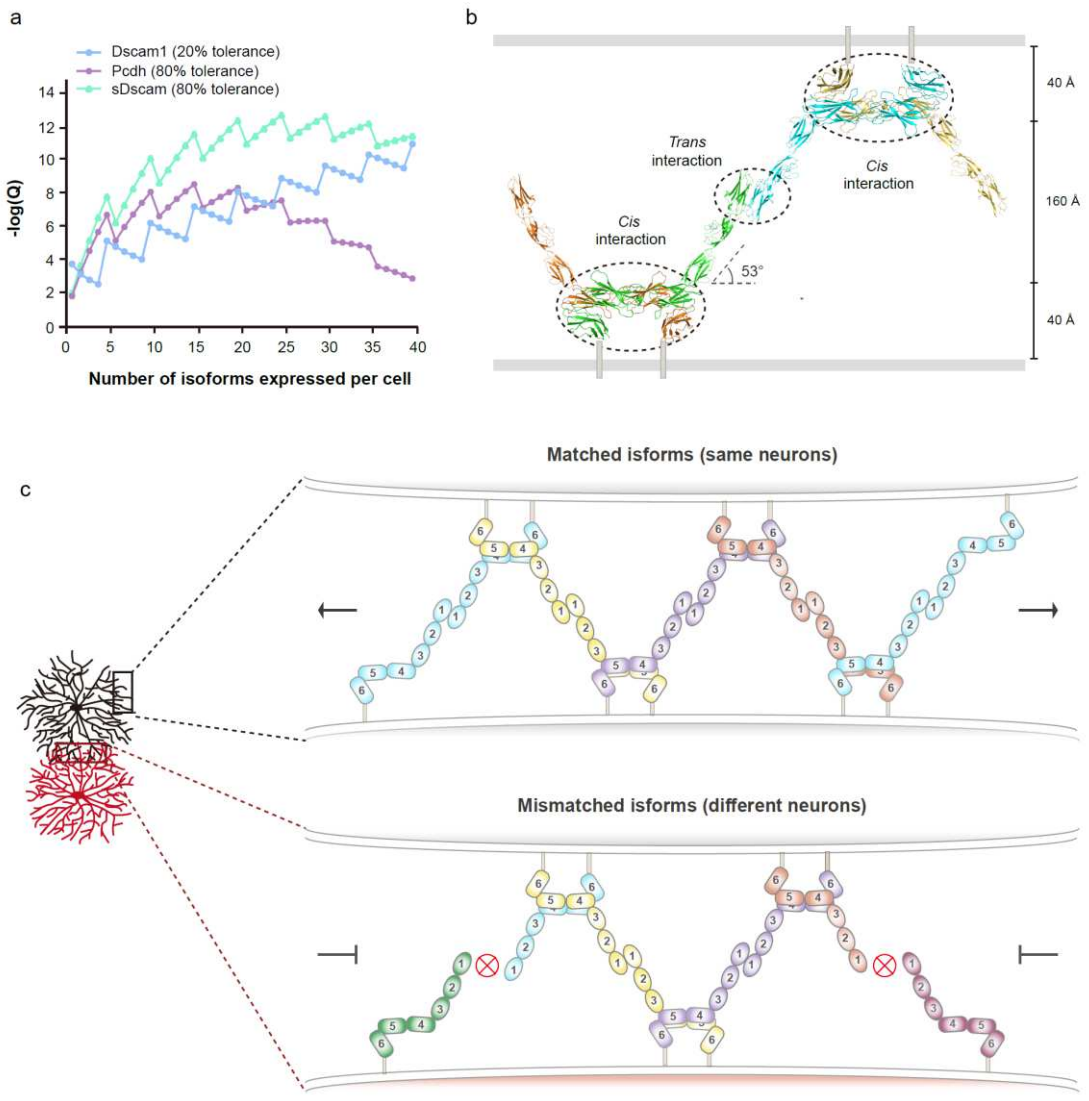


Figure 7. The zipper-like model for sDscam recognition. (**a**) Mathematical modeling of Dscam1-, Pcdh- or sDscam-mediated cell discrimination. Probabilities of incorrect non-self recognition between two contacting cells (shown on a logarithmic scale) are plotted as a function of the number of isoforms expressed per cell. (**b**) Schematic of the zipper-like structure of sDscam molecules between apposed membrane surfaces. The domains involved in the *cis* or *trans* interactions are indicated by dashed circles. (**c**) Proposed zipper-like model for sDscam recognition. Chain extension of sDscam molecules occurs between the dendrites of the same neuron which have the same isoform repertoire, resulting in the sDscam-mediated cell-cell recognition. On the contrary, dendrites of different neurons owning mismatching isoforms will terminate chain extension.

Supplementary

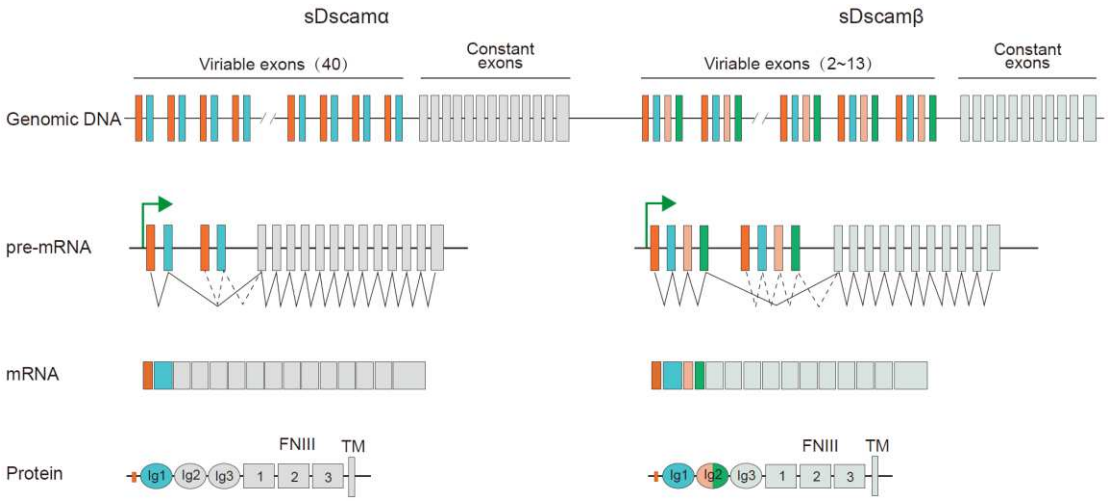


Figure S1. Organization of the two sDscam gene subfamilies in *M. martensii*. One α and 6 β subfamilies are composed of multiple tandemly arrayed regions (colored boxes) and common regions (grey boxes). Each sDscamα variable cassette encodes an N-terminal fragment and Ig1 domain, while that of sDscamβ encodes an N-terminal fragment and Ig1–Ig2 domains.

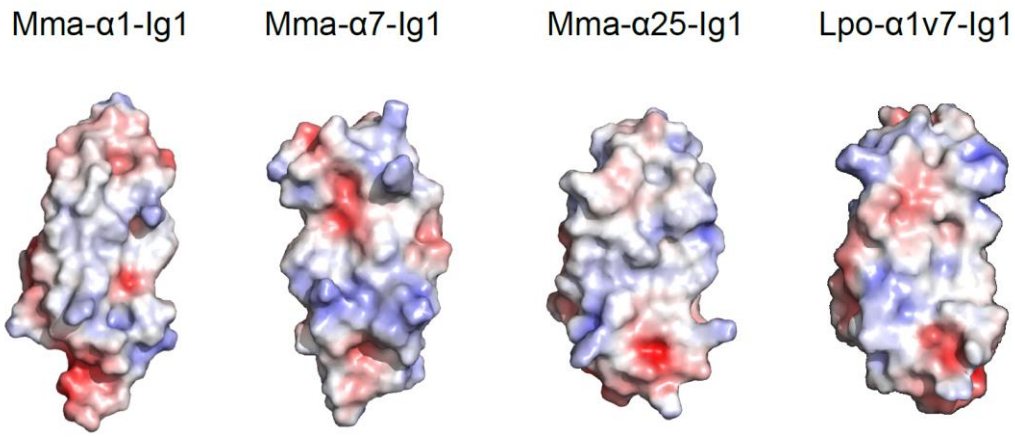


Figure S2. The electrostatic surface potential of the ABED face of sDscamα Ig1 isoforms. Blue and red (±5 kT/e) indicate the positively and negatively charged areas of the protein, respectively.

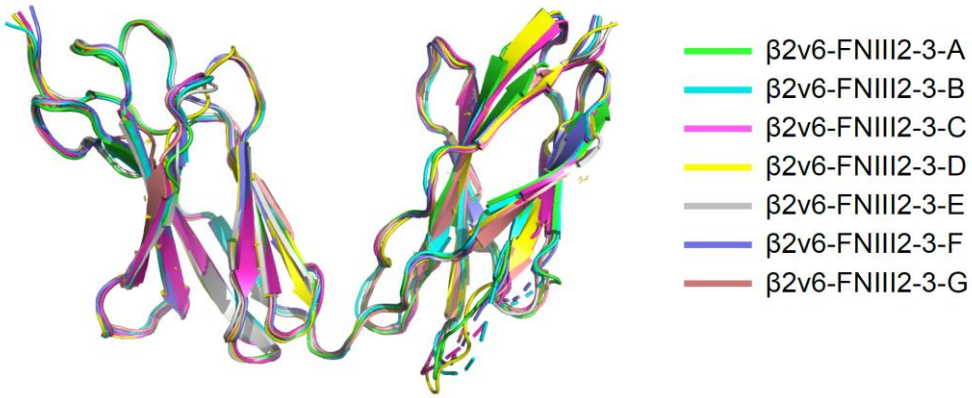
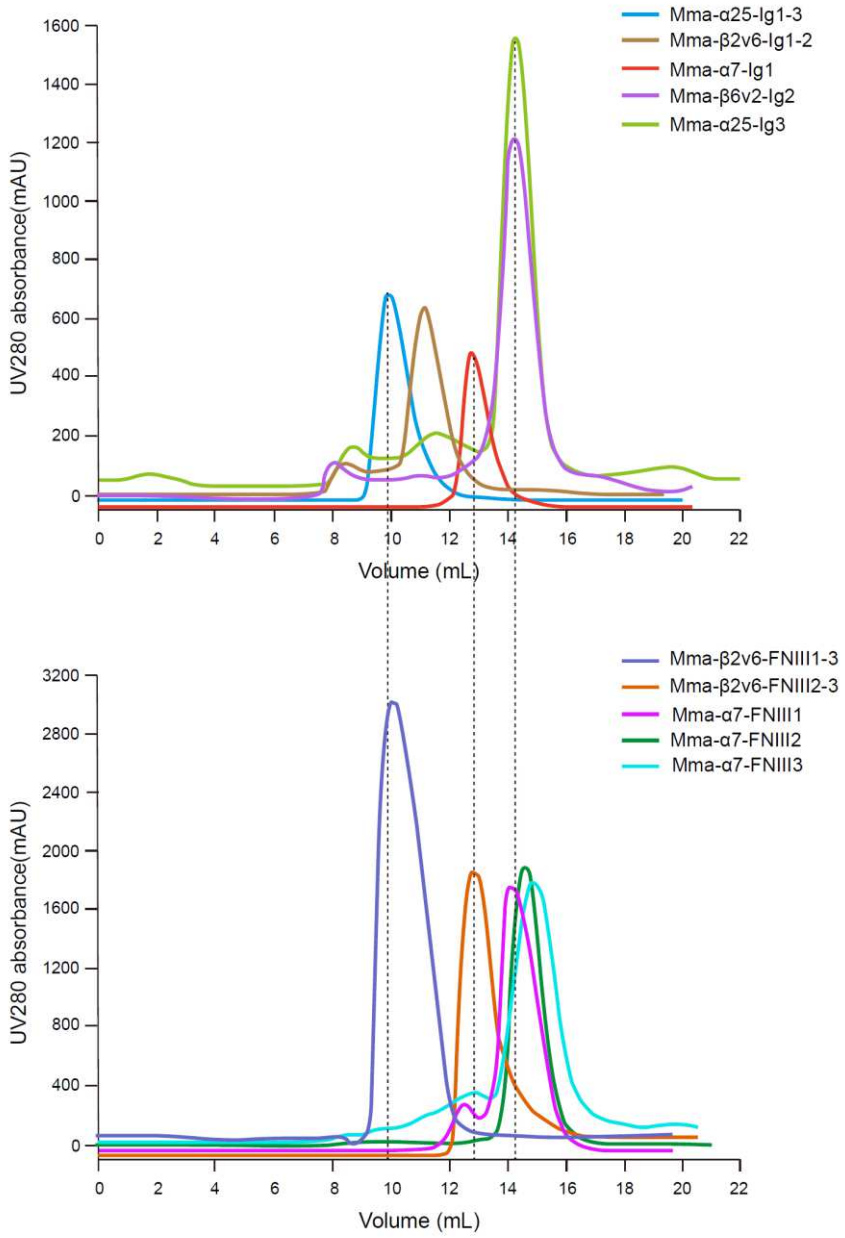


Figure S3. Superposition of sDscam FNIII2-3 fragments. The seven molecules per asymmetric unit in the FNIII2-3 crystal structure of sDscam isoform β2v6 are superimposed, which reveals an almost identical conformation.



606
607 **Figure S4. Gel-filtration profiles of different sDscam fragments.** Upper panel, the gel-
608 filtration profiles of Ig fragments. Lower panel, the gel-filtration profiles of FNIII fragments.
609 Dashed lines indicate the same elution volumes.
610

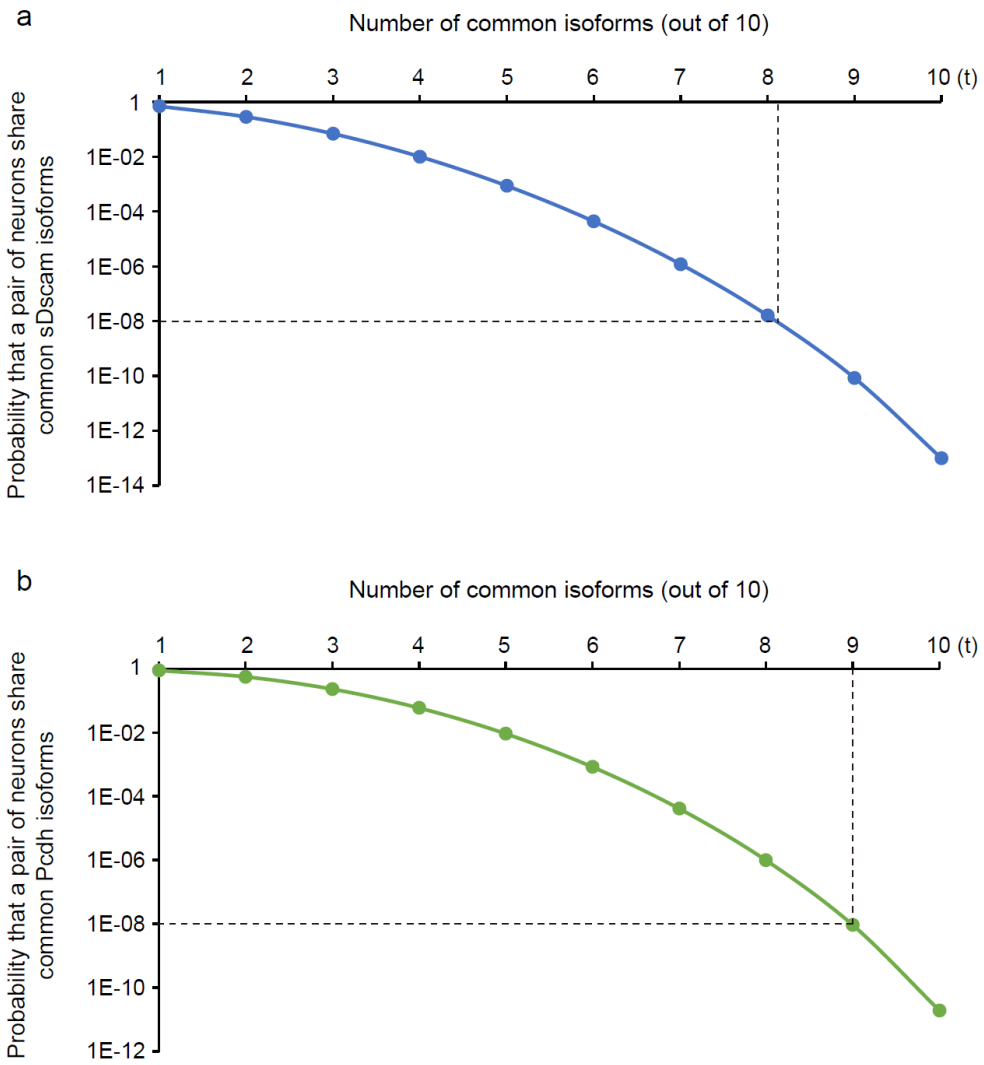


Figure S5. Mathematical modeling of random common isoforms. Probabilities that a pair of neurons share *t* or more isoforms in common out of 10 randomly expressed isoforms from 95 alternate sDscams (**a**) or 58 alternate Pcdhs (**b**) are calculated by Monte-Carlo simulations.

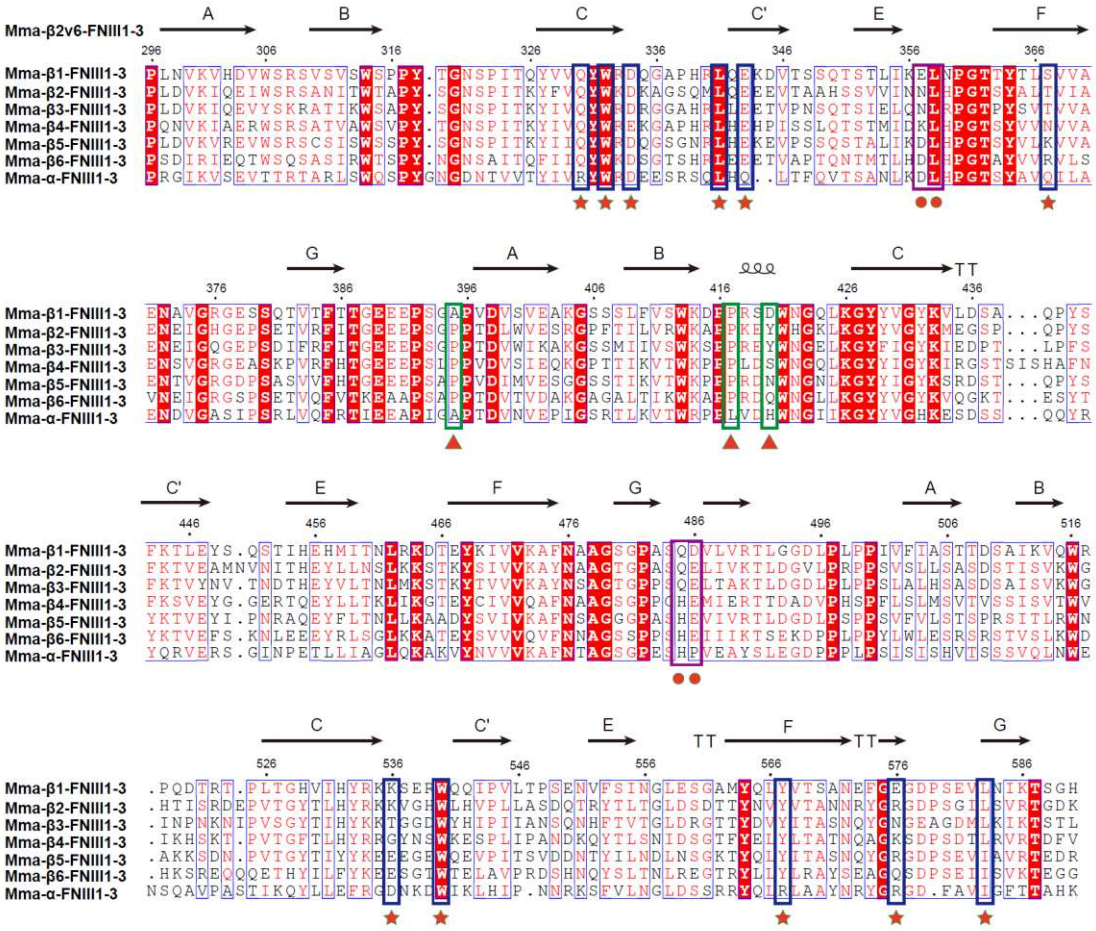


Figure S6. Sequence alignment of sDscam FNIII domains. The alignment is generated with CLUSTAL_X [32] and illustrated with ESPript [33]. The secondary structural elements and residue number are marked in accordance with sDscam isoform β2v6. Stars, filled circles and triangles below the sequences indicate the residues involved in the interactions of FNIII1-FNIII3, FNIII1-FNIII2 and FNIII2-FNIII2 interfaces, respectively.

625

626 **Table S1. Crystallization conditions.**

Proteins	Crystallization conditions
Mma-sDscamα 1-Ig1	40% (v/v) 2-Propanol, 100 mM Imidazole/Hydrochloric acid pH 6.5, 15% (w/v) PEG 8,000
Mma-sDscam α7-Ig1	0.2 M Ammonium acetate, 0.1 M Sodium citrate tribasic dihydrate pH 6.0, 25% (w/v) PEG 3,350, 5% (v/v) Jeffamine M-600 ® pH 7.0
Mma-sDscam β6v2-Ig1	0.2 M Magnesium Formate, 20% (w/v) PEG3,350
Mma-sDscam β2v6-Ig1-2	0.2 M magnesium chloride, 0.1 M HEPES pH7.0, 20 % (w/v) PEG 6,000
Mma-sDscam α25-Ig1-3	0.1 M KSCN, 15% (w/v) PEG 2,000, 0.1 M Glycine, 0.1 M Bis-Tris pH6.0
Lpo-sDscam α1v7-Ig1	1 M Sodium citrate tribasic, 100 mM Sodium cacodylate/ Hydrochloric acid pH6.5
Lpo-sDscam β3v7-Ig1	0.2 M $MgCl_2$, 0.1 M TRIS pH8.5, 20% (w/v) PEG 4,000
Mma-sDscam α7-FNⅢ1	25% (w/v) PEG 1,500, 100 mM SPG buffer pH 5.5
Mma-sDscam α7-FNⅢ2	0.1 M HEPES sodium pH 7.5, 1.5 M Lithium sulfate monohydrate
Mma-sDscam α7-FNⅢ3	0.1 M NaH_2PO_4 pH 6.5, 18% (w/v) PEG 8,000, 3% isopropanol
Mma-sDscam β2v6-FNⅢ23	0.2 M NaCl, 14% (w/v) PEG 8,000, 100 mM HEPES buffer pH 7.0
Mma-sDscam β2v6-FNⅢ123	20% (w/v) PEG 3,350, 100 mM PCB buffer pH 4.5

627

628

Table S2. Data collection and refinement statistics.

Construct	α1-Ig1	α7-Ig1	β2v6-Ig1-2	β6v2-Ig1	α25-Ig1-3	α1v7-Ig1	β3v7-Ig1	α7-FNIII1	α7-FNIII2	α7-FNIII3	β2v6-FNIII2-3	β2v6-FNIII1-3
PDB ID	7Y54	7Y4X	7Y9A	7Y95	7Y6O	7Y5J	7Y73	7Y8H	7Y5R	7Y8I	7Y6E	7Y8S
Data collection												
Space group	*C*2	$P4_32_12$	*I* 2	$P2_1$	*H*32	$P6_522$	$P2_1$	*C*2	$P2_12_12_1$	$P2_12_12_1$	*C*2	$P6_422$
Cell dimensions												
a, *b*, *c* (Å)	67.4, 40.7, 41.5	92.5, 92.5, 69.9	63.3, 49.0, 82.0	45.2, 39.1, 60.9	117.5, 117.5, 171.5	64.8, 64.8, 113.1	39.6, 40.9, 54.2	81.9, 43.2, 48.3	28.9, 53.3, 60.8	69.1, 85.8, 106.2	307.0, 56.2, 94.8	171.6, 171.6, 114.9
α, β, γ (°)	90.0, 94.2, 90.0,	90.0, 90.0, 90.0	90.0, 104.7, 90.0	90.0, 108.9 90.0	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 100.1, 90.0	90.0, 111.3, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 105.4, 90.0	90.0, 90.0, 120.0
Resolution (Å)	50.00 -1.80 (1.83-1.80)*	50.00-2.95 (3.00-2.95)	50.00-2.50 (2.54-2.50)	50.00-1.55 (1.58-1.55)	50.00-3.10 (3.17-3.10)	50.00-1.60 (1.63-1.60)	50.00-1.32 (1.36-1.32)	50.0-1.80 (1.84-1.80)	50.00-1.56 (1.60-1.56)	50.00-1.90 (1.93-1.90)	50.00-3.05 (3.10-3.05)	50.00-2.70 (2.75-2.70)
R_{pim} (%)	8.3 (11.9)	4.3 (34.5)	3.8 (9.9)	3.8 (32.7)	3.0 (26.1)	1.5 (17.1)	4.9 (11.8)	5.5 (28.4)	5.1 (13.4)	3.4 (38.1)	7.2 (43.3)	3.7 (38.2)
$I/\sigma I$	8.6 (6.5)	20.5 (2.0)	30.1 (6.9)	22.5 (1.7)	23.8 (2.0)	46.7 (4.0)	17.9 (1.8)	13.1 (2.2)	13.9 (10.0)	22.9 (2.0)	10.2 (1.7)	26.0 (2.0)
CC (1/2)	0.96 (0.94)	0.96 (0.74)	0.97 (0.97)	0.99 (0.82)	0.98 (0.90)	0.99 (0.94)	0.98 (0.87)	0.97 (0.78)	0.97 (0.95)	0.99 (0.67)	0.99 (0.72)	0.98 (0.79)
Completeness (%)	100.0 (99.8)	99.9 (99.4)	98.0 (96.6)	99.6 (97.9)	100.0 (99.6)	100.0(100.0)	98.1 (87.7)	99.3 (96.7)	100.0 (100.0)	99.8 (98.7)	98.7 (90.9)	100.0(100.0)
Redundancy	6.4 (6.1)	24.3 (17.4)	6.0 (5.1)	5.7 (4.1)	16.0 (12.2)	37.8 (35.9)	4.1 (2.7)	6.2 (4.2)	11.9 (9.7)	12.9 (10.7)	4.7 (3.8)	38.3 (31.8)
Refinement												
Unique reflections	10,339	6,660	8,285	30,014	7,501	19,192	35,491	14,018	13,650	46,459	25,974	25,039
R_{work}/ R_{free}	18.0/22.6	24.5/28.9	25.5/31.4	15.0/19.1	28.3/33.2	18.9/21.4	18.8/22.1	22.3/27.2	13.9/15.9	19.8/22.9	23.3/28.7	22.6/28.4
Ramachandran												
Favored (%)	96.94	95.74	94.68	98.45	90.38	98.92	99.47	95.43	99.03	98.23	93.11	96.72
Outlier (%)	0.00	0.00	0.00	0.00	0.34	0.00	0.00	0.57	0.00	0.00	0.00	0.00
R.m.s deviations												
Bond lengths (Å)	0.007	0.005	0.005	0.006	0.005	0.006	0.005	0.007	0.005	0.007	0.005	0.005
Bond angles (º)	0.745	1.390	0.921	0.718	0.578	0.906	0.786	0.759	0.738	0.862	1.033	0.847

*Highest resolution shell is shown in parenthesis.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- reportingSummary.pdf
- 7Y4XvalreportfullP1.pdf
- 7Y54valreportfullP1.pdf
- 7Y5JvalreportfullP1.pdf
- 7Y5RvalreportfullP1.pdf
- 7Y6EvalreportfullP1.pdf
- 7Y6OvalreportfullP1.pdf
- 7Y73valreportfullP1.pdf
- 7Y8HvalreportfullP1.pdf
- 7Y8IvalreportfullP1.pdf
- 7Y8SvalreportfullP1.pdf
- 7Y95valreportfullP1.pdf
- 7Y9AvalreportfullP1.pdf