Structural basis of Dscam isoform specificity

Rob Meijers[1,3]*†, Roland Puettmann-Holgado[2,4,8]*, Georgios Skiniotis[5], Jin-huan Liu[1,3], Thomas Walz[5], Jia-huai Wang[1,6,7] & Dietmar Schmucker[2,4]

The *Dscam* gene gives rise to thousands of diverse cell surface receptors[1] thought to provide homophilic and heterophilic recognition specificity for neuronal wiring[2–4] and immune responses[5]. Mutually exclusive splicing allows for the generation of sequence variability in three immunoglobulin ecto-domains, D2, D3 and D7. We report X-ray structures of the amino-terminal four immunoglobulin domains (D1–D4) of two distinct Dscam isoforms. The structures reveal a horseshoe configuration, with variable residues of D2 and D3 constituting two independent surface epitopes on either side of the receptor. Both isoforms engage in homo-dimerization coupling variable domain D2 with D2, and D3 with D3. These interactions involve symmetric, antiparallel pairing of identical peptide segments from epitope I that are unique to each isoform. Structure-guided mutagenesis and swapping of peptide segments confirm that epitope I, but not epitope II, confers homophilic binding specificity of full-length Dscam receptors. Phylogenetic analysis shows strong selection of matching peptide sequences only for epitope I. We propose that peptide complementarity of variable residues in epitope I of Dscam is essential for homophilic binding specificity.

The *Drosophila melanogaster* Dscam protein is an immunoglobulin superfamily (IgSF) member and consists of 10 immunoglobulin-like domains, 6 type III fibronectin domains, a transmembrane segment and a 374-residue cytoplasmic domain (Supplementary Fig. 1). Mutually exclusive alternative splicing occurs for exons 4, 6, 9 and 17, which encode the N-terminal half of Ig2 (D2), the N-terminal half of D3, the entire D7, and the transmembrane segment, respectively. In *Drosophila*, the combinatorial use of alternative exons potentially gives rise to 19,008 distinct extracellular receptor parts of Dscam[1]. Dscam was first identified as a highly diverse surface receptor required for neuronal wiring[1,6–13]. It was proposed that differences in isoforms expressed on the surface of neighbouring axons and dendrites determine their interactions[2,4,8]. Recent studies have demonstrated that Dscam is also required in the innate immune system[5,14]. Moreover, *in vitro* binding studies have shown that Dscam isoforms can interact in a highly selective homophilic manner, and that even closely related isoforms exhibit almost exclusive isoform-specific binding requiring the first eight immunoglobulin domains[2]. It was also proposed that modular interactions of the variable domains as D2^A^–D2^B^, D3^A^–D3^B^ and D7^A^–D7^B^ together are required to stabilize otherwise weak individual interactions[2].

The topology of the N-terminal eight immunoglobulin domains of Dscam was examined by negative-stain electron microscopy[15]. Averaged images of several isoforms reveal multiple distinct configurations indicative of remarkable flexibility (Fig. 1a, and Supplementary Fig. 2). In contrast, class averages of proteins of the N-terminal four immunoglobulin domains (D1–D4) revealed a horseshoe shape

for more than 90% of the molecules observed (Fig. 1b, and Supplementary Fig. 3).

D1–D4 proteins of two distinct Dscam isoforms were expressed using a baculovirus system and taken for crystallographic analysis (see Methods). The first isoform contains splicing variant 1 of exon 4 and splicing variant 34 of exon 6, designated as D1–D4$_{1.34}$. It crystallized in three forms, the best of which, $P4_222$, diffracted to 2.0 Å resolution. The 388-residue structure was determined using multiple anomalous dispersion and crystal averaging (Supplementary Table 1). It reveals a horseshoe-shaped configuration with D2 contacting D3 on the top and D1 contacting D4 at the bottom (Fig. 1c). The second isoform D1–D4$_{9.9}$ contains splicing variant 9 of exon 4 and variant 9 of exon 6. Its 3 Å resolution 391-residue structure in $C2$ form was obtained by molecular replacement. In spite of marked sequence differences in D2 and D3 (Supplementary Fig. 4), D1–D4$_{9.9}$ has a similar horseshoe configuration. Unless otherwise stated, D1–D4$_{1.34}$ is used for a description of structural features.

The horseshoe configuration is facilitated by a flexible 5-residue linker between D2 and D3 (Fig. 1c) and buries 2,550 Å^2 of surface area in the Dscam D1–D4$_{1.34}$ structure, a size that probably renders the configuration stable[16]. The interactions that stabilize the D2–D3 interface are mainly from conserved residues (Supplementary Fig 5a). Because D1 and D4 are constant domains, the D1–D4 interface is also probably conserved among all isoforms. Therefore the horseshoe configuration should be a general structural feature of all Dscam isoforms. This overall horseshoe topology is similar to the N-terminal four-domain structures of the insect protein hemolin[17] and the chicken neuronal receptor axonin[18] (Supplementary Figs 5b and 6). The prediction is that this configuration might represent a common feature for other neural cell adhesion molecules of the L1 family[19].
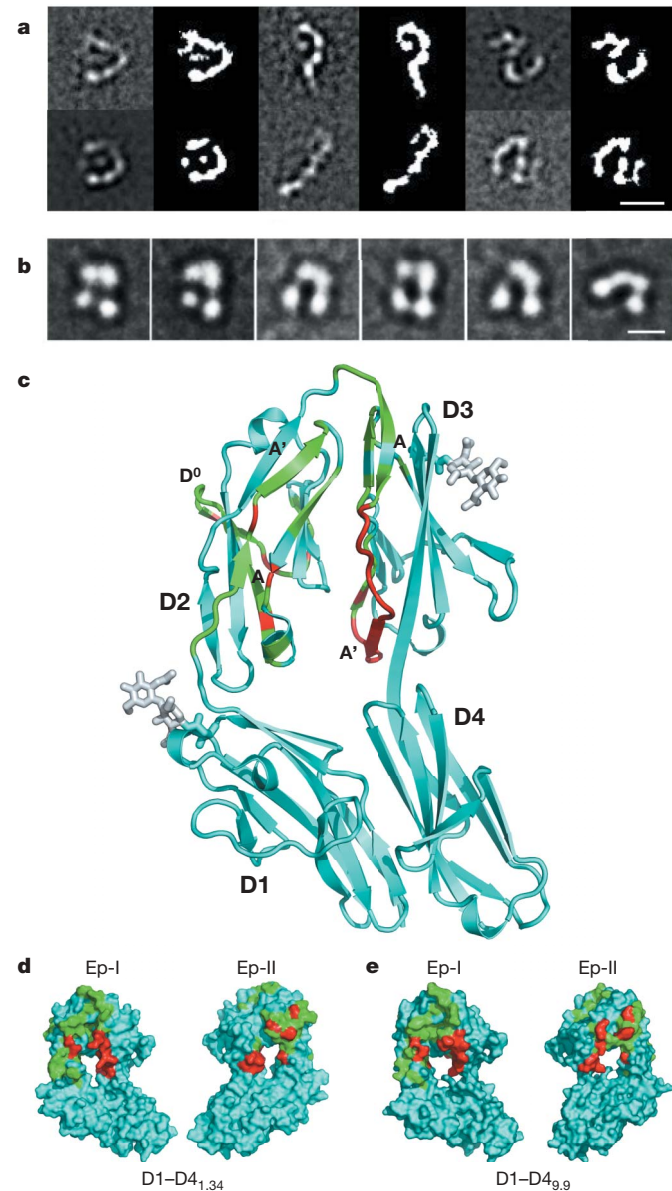
The four Dscam N-terminal domains most closely resemble the I-set of the IgSF fold[20]. The alternatively spliced segment of D2 (Supplementary Fig. 4) exhibits two features that are unique to Dscam. First, both isoforms contain an unusually long A' strand (Fig. 1c), and, second, an extra β-strand (termed D^0) that emerges after the C' strand, folding alongside the D strand (Fig. 1c, and Supplementary Fig. 7a, b). This C'–D^0 unit encompasses the most variable region of exon 4 (Supplementary Fig. 4). The alternatively spliced segment in D3 contains an unusually elongated transition of the A–A'. In D1–D4$_{1.34}$, this prominent 8-residue A–A' bulges out (Fig. 1c, and Supplementary 7c), whereas in the D1–D4$_{9.9}$, this 7-residue A–A' curves into a single 3_{10} helical turn (Supplementary Fig. 7d). The long A' strand of D2 and the distinct A–A' protrusion of D3 are exposed on the same front face and, as a consequence of the horseshoe configuration, assemble into a composite surface epitope. It displays variable residues that we define and refer to as epitope I

[1]Department of Medical Oncology and [2]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. [3]Department of Medicine, [4]Department of Neurobiology, [5]Department of Cell Biology, [6]Department of Pediatrics and [7]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA. [8]Institute for Biology III, University of Freiburg, D-79104 Freiburg im Breisgau, Germany. †Present address: Synchrotron Soleil, L'Orme des Merisiers, 91192 Saint Aubin, France.
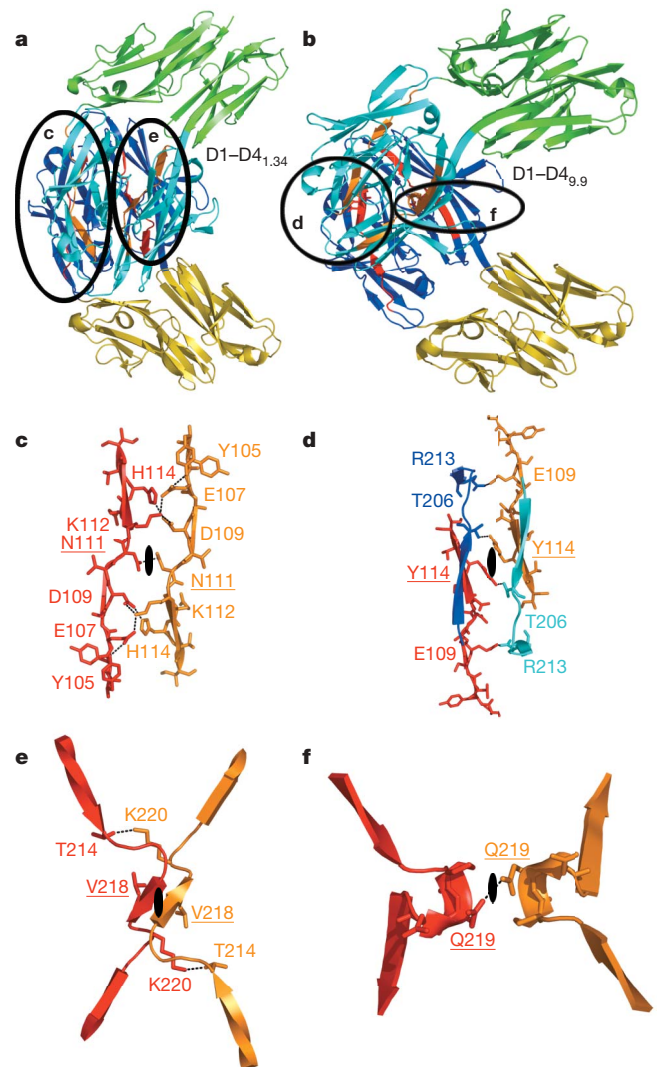*These authors contributed equally to this work.

(Fig. 1d, e). The C'–D^0 unit of D2 and the β-strand B of D3 reside on the opposite face, constituting an entirely separate epitope, which we define and refer to as epitope II (Fig. 1d, e). Both exon 4 and exon 6 contribute to each epitope in such a manner that their N-terminal segments encode epitope I and the C-terminal segments encode epitope II sequences.

Molecular packing analyses of all crystal forms of both isoforms show that epitope I is the key interaction surface engaged in

homophilic dimerization (Fig. 2). The central structural element in the D2^A–D2^B interactions between the monomers A and B is the long and highly variable A' strand. In D1–D4$_{1.34}$, this segment is EADVNKEH, starting at residue 107. A hydrogen-bonding network is formed between peptide segments Glu 107 to His 114 of interacting D2s (Fig. 2a, c). At the centre of the dimer are two side-chain hydrogen bonds formed by a pair of symmetry-related residues (Asn 111). There are two salt bridges (Glu 107^A–Lys 112^B and Asp 109^A–Lys 112^B) as well as a hydrogen bond (Asp 109^A–His 114^B) from molecules A to B, and vice versa. In D1–D4$_{9.9}$, the equivalent segment, starting at residue 109, is ESEADNEY. The phenol rings of dyad-related Tyr 114 pack against each other (Fig. 2b, d) at the centre of the D2^A–D2^B interface. The Tyr 114 side chain of one molecule is anchored by a hydrogen bond with the side chain of the conserved residue Thr 206 of a dyad-related molecule.

Figure 1 | Structure of the N-terminal four-domain fragment of Dscam.
a, Representative class averages from negatively stained Dscam D1–D8$_{1.34.30}$ show that the molecule can adopt different conformations but retains the horseshoe configuration of the N-terminal D1–D4 domains. Scale bar, 10 nm. **b**, Representative class averages from negatively stained Dscam D1–D4$_{1.34}$ show that the four domains of Dscam D1–D4$_{1.34}$ are arranged in a horseshoe configuration. Scale bar, 5 nm. **c**, Ribbon diagram of Dscam D1–D4$_{1.34}$ coloured according to sequence variability; conserved residues are coloured cyan, variable residues are green and hypervariable residues are red. The variability was calculated using Shannon's uncertainty[22], and residues were classified as hypervariable if the uncertainty value exceeded two-thirds of the highest value observed for all residues from exons 4 and 6. **d**, **e**, Surface representation of epitope I (left) and II (right) on either side of the horseshoe for Dscam D1–D4$_{1.34}$ (**d**) and Dscam D1–D4$_{9.9}$ (**e**). Colour codes are as in **c**. The figure was prepared using PyMOL (http://www.pymol.org).

Figure 2 | Homophilic dimers observed in the crystal lattice. a, b, Ribbon diagram of the dimer in Dscam D1–D4$_{1.34}$ (**a**) and Dscam D1–D4$_{9.9}$ (**b**). D1 and D4, green; D2 and D3, blue for monomer A; D1 and D4, yellow; D2 and D3, cyan for monomer B. Residues at symmetry centre are underlined. The isoform-specific interaction elements are shown as red and orange in molecules A and B, respectively, and are displayed in more detail along their respective twofold axes: **c**, the D2^A–D2^B interface of Dscam D1–D4$_{1.34}$; **d**, the D2^A–D2^B interface of Dscam D1–D4$_{9.9}$ (blue and cyan residues are constant); **e**, the D3^A–D3^B interface of Dscam D1–D4$_{1.34}$; **f**, the D3^A–D3^B interface of Dscam D1–D4$_{9.9}$. Residues involved in dimer-sustaining hydrogen bonds are labelled and the dyad axes are displayed as black ellipsoids. The figure was prepared using PyMOL (http://www.pymol.org).

The central structural element in the $D3^A$–$D3^B$ interface is the highly variable A–A' protrusion. In D1–D4$_{1.34}$, the isoform-specific sequence for this segment, starting at residue 214, is TPALVQKP. From Ala 216 to Lys 220, the main chains of the two protomers pair into an anti-parallel mini β-sheet. The interaction is further stabilized by a symmetry-pair of specific hydrogen bonds (Lys 220^A–Thr 214^B). A pseudo-dyad runs through the middle of this β-sheet, where two symmetry-related Val 218 residues touch (Fig. 2e). In the D1–D4$_{9.9}$ dimer, there are, again, interactions between these opposing A–A' protrusions, with a sequence of VNPQDKH starting at residue 216, but of a different nature. The most exposed residues Pro 218 and Gln 219 at the 3_{10} helical turn's acme contribute prominent contacts to the dimer interface (Fig. 2f). One hydrogen bond is formed between side chains of a symmetry-related pair of Gln 219. In both isoforms, we noticed only minor cross interactions between D2 and D3, namely a salt bridge between the variable Arg 213 of D3 and the conserved Asp 133 of D2 in D1–D4$_{1.34}$, and a similar pair between Arg 213 of D3 and Glu 109 of D2 (both variable residues) in D1–D4$_{9.9}$.

Three key features characterize the homophilic dimerization of both isoforms: (1) the most variable residues of epitope I but not epitope II constitute the core of the dimer interface; (2) in a similar manner to other immunoglobulin-domain-containing adhesion molecules[21], the interface is dominated by hydrophilic residues, and the binding strength is expected to be low; and (3) within the dimer there are separable $D2^A$–$D2^B$ and $D3^A$–$D3^B$ interfaces between molecules A and B. These shared features are predicted to be applicable to all 576 (12 D2 × 48 D3) potential homo-dimer interactions. Although the central contact area of dimerization is clearly defined by the variable residues of epitope I in the two isoform structures, their orientation and the significantly different extent of buried surface area (3,711 Å^2 for the D1–D4$_{1.34}$ dimer and 1,220 Å^2 for the D1–D4$_{9.9}$ dimer) indicate a considerable diversity in the orientation and binding strength of homophilic isoform pairs.

Biochemical experiments indicate that Dscam homophilic binding is remarkably robust. However, homophilic interactions were only observed if the isoforms are nearly identical and contain the first eight immunoglobulin domains[2]. Therefore, the question arises as to how much the local intermolecular contacts of D2 and D3 in D1–D4 homodimers (Fig. 2) contribute to the specificity of full-length Dscam homodimers. To address this, we introduced point mutations in residues critical to $D2^A$–$D2^B$ and $D3^A$–$D3^B$ interactions. Homophilic binding was tested using a previously described bead-to-cell adhesion assay[2]. Beads that contain the entire extracellular part of the Dscam$_{1.34.30}$ isoform bind efficiently to COS cells expressing the identical Dscam$_{1.34.30}$ receptor but not at all to cells expressing Dscam$_{1.30.30}$ (Fig. 3a, and Supplementary Figs 8 and 9).

In D1–D4$_{1.34}$, the $D2^A$–$D2^B$ interface consists of the A' strand from residues Glu 107 to His 114. Therefore, K112 E and H114D mutations will probably disrupt the hydrogen bonding in the interface. Indeed, the bead-to-cell binding assay shows that this double mutation reduces homophilic binding to only 4% compared to wild type (Fig. 3). The $D3^A$–$D3^B$ interface is marked by the formation of a mini anti-parallel β-sheet, involving four main chain hydrogen bonds centred at Val 218. To disrupt this mini β-sheet, we mutated Val 218 to a proline, which reduced the homophilic binding to only 14% (Fig. 3a, and Supplementary Fig. 8).

Cell adhesion often relies on highly specific multivalent interactions, composed of relatively weak single interaction pairs[21], which also seems to be the case for homophilic binding of Dscam. We mutated two residues within epitope I in tandem to alanine to reduce specific inter-domain hydrogen bonds. These mutations had a moderate effect on the cell adhesion properties of Dscam, reducing binding to 68% for N111A/K112A of D2 and 74% for L217A/Q219A of D3 (Fig. 3a, and Supplementary Fig. 8). When we combine tandem mutants for both domains, the reduction in adhesion is much more pronounced, with 31% for N111A/K112A/K220A/P221A and 10%
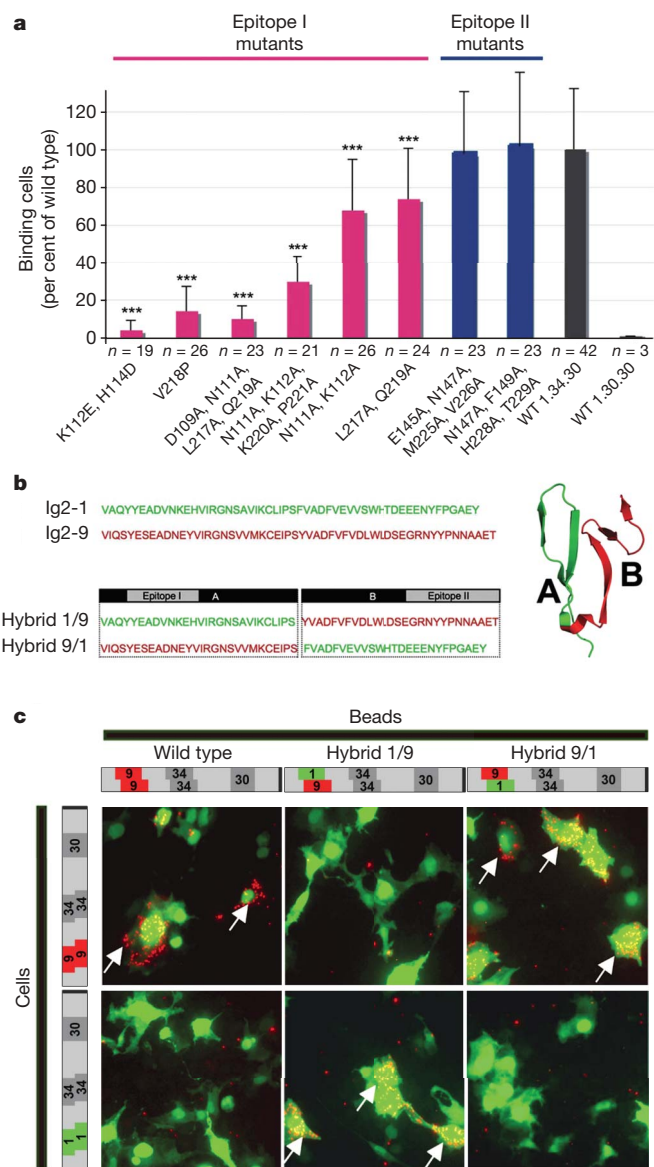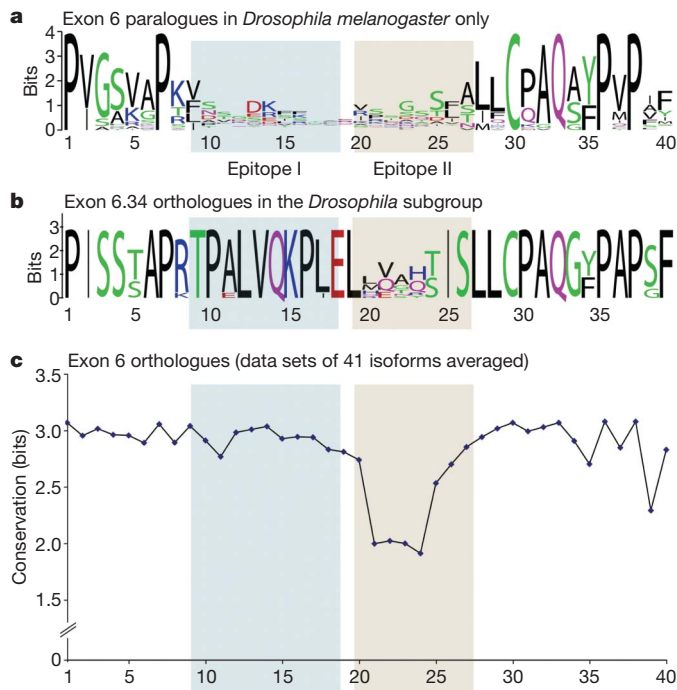

Figure 3 | Epitope I confers homophilic binding specificity.
a, Quantification of binding of wild-type Dscam receptor and eight different isoform variants bearing mutations in either epitope I or epitope II to Cos cells expressing wild-type Dscam$_{1.34.30}$. Representative micrographs depicting bead-to-cell binding are given in Supplementary Figs 9 and 10. The total numbers of counted fields, each containing 50–200 cells, are given below each bar. Error bars indicate standard deviation (as variation between entire optical fields (10 × 10 microscope magnification)). ***$P > 0.005$, t-test. **b**, Peptide sequences for the two different Ig2 domains containing exon 4.1 or 4.9 (Ig2-1 or Ig2-9, respectively). Hybrid isoforms 1/9 and 9/1 were generated by overlap-PCR from exon 4 templates. Hybrid 1/9 incorporates the N-terminal half of exon 4 (segment A), including epitope I sequences from isoform 4.1, and the C-terminal half of exon 4 (segment B), including epitope II sequences from isoform 4.9. Hybrid 9/1 incorporates the equivalent 4.9 and 4.1 sequences. The structural model shows the distinct spatial separation of peptide segments. **c**, Bead-to-cell binding assays were performed (see Supplementary Figs 9 and 10) using Dscam protein with either wild-type Ig2-9 and Ig2-1 sequences (left panels) or hybrid sequences (middle and right panels for hybrid design see **b**). The schematics next to the micrographs indicate exon combinations (exon 4 in red or green boxes) of the proteins tested. Both hybrid proteins fully retained their homophilic binding capabilities as indicated by robust binding of beads to cells (arrows). Full binding was observed only when epitope I sequences were matching. Homophilic binding of beads to cells was indistinguishable from 'donor' wild-type isoforms, even if epitope II sequences were completely different. In contrast, binding was completely abolished with non-matching epitope I but matching epitope II sequences.

for D109A/N111A/L217A/Q219A. We also mutated several surface-exposed residues of epitope II. In this case, binding assays revealed no difference in adhesion between wild type and the two quadruple mutant isoforms E145A/N147A/M225A/V226A and N147A/F149A/H228A/T229A (Fig. 3a, and Supplementary Fig. 8).

To confirm further that epitope I determines binding specificity, we generated hybrid receptor isoforms swapping either the N-terminal segment of D2 (segment A, containing epitope I; Fig. 3b) or the C-terminal segment of D2 (segment B, containing epitope II; Fig. 3b). We found that hybrid 1/9 (exon 4.1 N terminus; exon 4.9 C terminus) can only bind to isoform $Dscam_{1.34.30}$ but not to isoform $Dscam_{9.34.30}$ (Fig. 3c). In contrast, hybrid 9/1 (exon 4.9 N terminus; exon 4.1 C terminus) will not bind to $Dscam_{1.34.30}$ but binds well to $Dscam_{9.34.30}$. Hybrid isoforms bearing identical epitope II but different epitope I sequences did not show binding to wild-type isoforms (Fig. 3c). Therefore, homophilic binding specificity in these hybrid receptors is determined by the epitope I sequence.

Because the presented structures document only 2 of the 576 possible exon 4/exon 6 combinations, we sought to potentially discern a general trend by using a phylogenetic comparison, expecting that the proposed functional differences between epitope I and epitope II sequences are reflected in differential sequence conservation. We analysed the sequence conservation[22] of epitope I and epitope II among *Drosophila* species and determined the variability among paralogues as well as orthologues (Fig. 4, and Supplementary Figs 11, 12). Alignments of 12 orthologues of the *Drosophila melanogaster* isoform 6.34 show high conservation for residues along the A–A' protrusion of D3 (epitope I sequence), but low conservation of residues along the B strand of D3 (epitope II sequence) (Fig. 4, and

Supplementary Fig. 10). In concurrence, orthologues of isoform 4.1 show high conservation for epitope I along the A' strand of D2, but low conservation for epitope II along the C'–D^0 unit of D2 (Supplementary Figs 11, 12). The sequences of epitope I segments potentially involved in homophilic interactions are, therefore, highly conserved between species, whereas the sequences that constitute epitope II are more divergent.

In summary, we have provided a structural analysis of the recognition specificity of two variable immunoglobulin domains of *Drosophila* Dscam. Although the D1–D4 structures reported here contain only two variable domains, and it remains to be determined how D7 contributes to binding, our biochemical analysis in the context of the full-length Dscam receptor (Fig. 3) is consistent with an essential contribution of the variable peptide segments of epitope I to the homophilic-binding specificity of Dscam. Swapping the peptide segment containing epitope I but not epitope II resulted in a full switch in binding specificity between two isoforms. This strongly suggests that in a Dscam dimer the matching epitope I peptides enable binding, and non-matching ones inhibit homophilic binding, thereby functioning as a specificity module. The strong sequence conservation of epitope I residues is consistent with a high evolutionary selection pressure preserving a limited set of homophilic-binding interfaces. Although an involvement of epitope II in binding of non-Dscam ligands has not been tested experimentally, the apparently faster-evolving sequence variability in epitope II would be consistent with immune receptor adaptations to dynamic alterations in host–pathogen interactions. We therefore hypothesize that this structural separation of homophilic and heterophillic binding (that is potentially self and non-self recognition) in Dscam may have enabled the parsimonious use of the same gene in creating a large receptor diversity in both the nervous system and immune system.

METHODS SUMMARY

Dscam constructs were expressed both in *Pichia pastoris* and the baculovirus expression system (Invitrogen). For electron microscopy, uranyl formate was used for negative staining, as described[15]. The X-ray structure of the D1–D4$_{1.34}$ construct was solved by multiple anomalous dispersion on a K_2PtCl_4 derivative at the X25 beamline of the National Synchrotron Light Source, and refined to a final R factor of 17.3% (R_{free} = 20.5%). The X-ray structure of the D1–D4$_{1.34}$ construct was solved by molecular replacement and refined to a final R factor of 27.1% (R_{free} = 30.3%). The sequence analysis is described in the Supplementary Information. The cloning and protein purification of the full-length extracellular mutant constructs of Dscam, as well as the bead aggregation assays, were done as described[2] with some modification described in detail in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Figure 4 | Phylogenetic comparison reveals differential sequence conservation of epitopes I and II. a, Sequence logo representation[22] of the conservation of exon 6 isoforms (paralogues) within *Drosophila melanogaster*. A total of 38 isoforms that encode exactly 40 amino acids were used for comparison. Bits, on the *y* axis, indicates units for evolutionary conservation[22]. **b,** Sequence logo representation of the conservation of 39 amino acids among orthologues of exon 6.34 in the *Drosophila* subgroup. **c,** Average interspecies conservation of amino acid sites in exon 6. The analysis includes exon 6 isoforms for which orthologies could be established. Average values for 41 data sets (each calculated from at least 8 aligned species) are plotted. Sites with the highest bit score are conserved between all species. Epitopes I and II are shaded in all panels

1. Schmucker, D. *et al.* *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* **101**, 671–684 (2000).
2. Wojtowicz, W. M., Flanagan, J. J., Millard, S. S., Zipursky, S. L. & Clemens, J. C. Alternative splicing of *Drosophila* Dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding. *Cell* **118**, 619–633 (2004).
3. Schmucker, D. & Flanagan, J. G. Generation of recognition diversity in the nervous system. *Neuron* **44**, 219–222 (2004).
4. Neves, G., Zucker, J., Daly, M. & Chess, A. Stochastic yet biased expression of multiple Dscam splice variants by individual cells. *Nature Genet.* **36**, 240–246 (2004).
5. Watson, F. L. *et al.* Extensive diversity of Ig-superfamily proteins in the immune system of insects. *Science* **309**, 1874–1878 (2005).
6. Wang, J., Zugates, C. T., Liang, I. H., Lee, C. H. & Lee, T. *Drosophila* Dscam is required for divergent segregation of sister branches and suppresses ectopic bifurcation of axons. *Neuron* **33**, 559–571 (2002).
7. Hummel, T. *et al.* Axonal targeting of olfactory receptor neurons in *Drosophila* is controlled by Dscam. *Neuron* **37**, 221–231 (2003).
8. Zhan, X. L. *et al.* Analysis of Dscam diversity in regulating axon guidance in *Drosophila* mushroom bodies. *Neuron* **43**, 673–686 (2004).
9. Zhu, H. *et al.* Dendritic patterning by Dscam and synaptic partner matching in the *Drosophila* antennal lobe. *Nature Neurosci.* **9**, 349–355 (2006).
10. Chen, B. E. *et al.* The molecular diversity of Dscam is functionally required for neuronal wiring specificity in *Drosophila*. *Cell* **125**, 607–620 (2006).

11. Hughes, M. *et al.* Homophilic Dscam interactions control complex dendrite morphogenesis. *Neuron* **54**, 417–427 (2007).
12. Matthews, B. *et al.* Dendrite self-avoidance is controlled by Dscam. *Cell* **129**, 593–604 (2007).
13. Soba, P. *et al. Drosophila* sensory neurons require Dscam for dendritic self-avoidance and proper dendritic field organization. *Neuron* **54**, 403–416 (2007).
14. Dong, Y., Taylor, H. E. & Dimopoulos, G. AgDscam, a hypervariable immunoglobulin domain-containing receptor of the *Anopheles gambiae* innate immune system. *PLoS Biol.* **4**, e229 (2006).
15. Ohi, M., Li, Y., Cheng, Y. & Walz, T. Negative staining and image classification— powerful tools in modern electron microscopy. *Biol. Proced. Online* **6**, 23–34 (2004).
16. Lo Conte, L., Chothia, C. & Janin, J. The atomic structure of protein–protein recognition sites. *J. Mol. Biol.* **285**, 2177–2198 (1999).
17. Su, X. D. *et al.* Crystal structure of hemolin: a horseshoe shape with implications for homophilic adhesion. *Science* **281**, 991–995 (1998).
18. Freigang, J. *et al.* 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 (2000).
19. Schurmann, G., Haspel, J., Grumet, M. & Erickson, H. P. Cell adhesion molecule L1 in folded (horseshoe) and extended conformations. *Mol. Biol. Cell* **12**, 1765–1773 (2001).
20. Harpaz, Y. & Chothia, C. Many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to a new structural set which is close to that containing variable domains. *J. Mol. Biol.* **238**, 528–539 (1994).
21. Wang, J. Protein recognition by cell surface receptors: physiological receptors versus virus interactions. *Trends Biochem. Sci.* **27**, 122–126 (2002).
22. Schneider, T. D. & Stephens, R. M. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097–6100 (1990).

METHODS

Cloning, expression and protein purification for structural studies. Nucleotide sequences encoding Dscam D1–D4$_{1.34}$ and D1–D4$_{9.9}$ were amplified from complementary DNA plasmids by PCR. Two plasmids containing combinations of alternative exons 4 and 6 (4.1&6.34 and 4.9&6.9) were used as templates. An N-terminal His-tag followed by a TEV-protease recognition site and *Sac*I and *Hind*III restriction sites were added. Recombinant baculovirus was generated by recombination with Bac-N-Blue DNA by Invitrogen. Virus was amplified in *Spodoptera frugiperda* (SF9) cells (Invitrogen) and High five cells to a density of 1.5×10^6 cells ml^{-1} and infected with the recombinant baculovirus at a multiplicity of infection between 1 and 10. At 72 h post infection, cells were spun down and the medium was filtered using a 0.22 µm CA membrane (Corning) and dialysed overnight against PBS, pH 7.0, supplemented with 100 mM NaCl. The medium was loaded on a Ni-NTA column (Qiagen) equilibrated with PBS, pH 7.0, 100 mM NaCl and 10 mM imidazole by gravity flow. Recombinant Dscam was eluted with PBS, pH 7.0, 100 mM NaCl and 200 mM imidazole. The His-tag was cleaved with TEV protease (Invitrogen) and subsequently removed by another run over the Ni-NTA column. Any remaining contaminating proteins were removed by gel filtration using a Superdex 200 column. The protein was concentrated to 20 mg ml^{-1} in 20 mM Hepes buffer, pH 7.4.

The inducible eukaryotic *Pichia pastoris* expression system (EasySelect, Invitrogen) was also used to produce the N-terminal-eight-domain constructs. Dscam D1–D8 constructs were cloned into pPCIZα (Invitrogen) with an N-terminal His(6)-tag followed by a TEV-protease recognition site. For transformation, electrocompetent *Pichia* cells were prepared according to a standard protocol (Invitrogen) and 6–15 µg of the expression plasmid, linearized with *Pme*I (NEB), was added to 80 µl of cells and incubated on ice for 10 min. Cells were electroporated using a Micropulser (Biorad) and cells were resuspended carefully in 1 ml of cold 1 M sorbitol solution. Cells were allowed to recover at 30 °C without shaking for 1 h and another 30 min after addition of 500 µl YPD-medium. Cell suspension (120 µl) was plated on YPD-agar plates containing 100 µg or 300 µg of zeocin, and plates were screened for colonies after 3–4 days.

For expression screening of clones, colonies were picked and resuspended into 4 ml of BMGY-medium in deep 24-well plates (Whatman) and grown for 24 h. Cells were allowed to settle and the medium was exchanged for 1 ml BMMY to induce expression. Medium was harvested and spun after 16–20 h and loaded on a 4–12% Bis-Tris gel under denaturing conditions. For large-scale protein production, strongly expressing clones were grown in 500 ml BMGY for 24 h, spun, and resuspended in 100 ml BMMY. After 20 h, cells were spun down, medium was filtered and the recombinant protein was purified using Ni-NTA (Qiagen) affinity-chromatography. After washing with 12 bed volumes of HEPES, pH 7.5, buffer was supplemented with 100 mM NaCl and 25 mM imidazole. Protein was eluted with HEPES, pH 7.5, 100 mM NaCl and 250 mM imidazole. Typically, yields were 30–60 mg l^{-1}. For removal of the N-terminal His-tag, the protein was incubated for 24 h at room temperature with recombinant TEV-protease (Invitrogen).

Crystallization and structure determination. Crystals of Dscam D1–D4$_{1.34}$ were grown in 1.5 M ammonium sulphate and 0.1 M Hepes buffer at pH 7.5. Three crystal forms were obtained (*P*4$_2$22: $a = b = 99.2$ Å, $c = 164.0$ Å; *I*4$_1$22: $a = b = 146.7$ Å, $c = 325.5$ Å; *C*222$_1$, including the His-tag: $a = 99.8$ Å, $b = 166.8$ Å, $c = 125.6$ Å). Crystals of Dscam D1–D4$_{9.9}$ were grown in 10% PEG 8000, 1 mM spermidine and 0.1 M Tris-HCl at pH 8.5 in space group *C*2 with cell dimensions of $a = 277.8$ Å, $b = 70.5$ Å, $c = 72.8$ Å and $\beta = 105.1°$. Native data were collected with a Quantum Q315 CCD detector (ADSC) at the 19ID beamline at the Advanced Photon Source. Multi-wavelength anomalous dispersion data were collected from the *I*4$_1$22 crystal form of D1–D4$_{1.34}$ with a Quantum Q315 CCD detector (ADSC) at the X25 beamline at the Brookhaven National Synchrotron Light Source. Three wavelengths were used (see Supplementary Table 1), and data were processed and scaled with HKL2000 (ref. 23). The Pt sites were located with SHELXD[24] using data from 12 to 3.7 Å. The phases were refined with SHARP[25] to a resolution of 3.5 Å with a Figure of Merit of 0.38. The phased map was solvent-flattened and averaged with DM[26] to 3.5 Å resolution along the translation vector (0.0, 0.0, 0.5) as obtained from the native Patterson. Iterative cycles of manual model building with Xtalview[27] and multi-crystal averaging with DMMULTI[28] were performed and model refinement was switched to the *P*4$_2$22 crystal form, extending to 2.0 Å resolution. Arpwarp[29] was used to further improve the map by iterative rebuilding (in the molrep mode) and subsequent tracing of the model. A total of 350 out of 388 residues were automatically traced, and 349 side chains docked. The model was completed manually and refined to an *R*-factor of 17.3% (*R*$_{free}$ 20.5%) with Refmac5 (ref. 30) using a bulk solvent correction, including hydrogens and TLS refinement. The final model contains 388 amino acid residues, two glycans and 709 water molecules. The Ramachandran statistics indicate 98.3%

of the residues are in the most-favoured region and 1.7% of the residues are in the generously allowed region. The Dscam D1–D4$_{9.9}$ structure was solved by molecular replacement with MOLREP[4] using Dscam D1–D4$_{1.34}$ as a search model. The structure was refined at 3.0 Å to an *R*-factor of 27.1% (*R*$_{free}$ 30.3%) with PHENIX[31] using non-crystallographic symmetry and B group refinement, and contains 391 residues. The Ramachandran statistics indicate 96.8% of the residues are in the most-favoured region and 3.2% of the residues are in the generously allowed region.

Electron microscopy and image processing. Dscam constructs D1–D4 and D1–D8 were prepared for electron microscopy using the conventional negative-staining protocol[32]. Briefly, 3 µl of sample were adsorbed to a glow-discharged carbon-coated copper grid and stained with 0.75% uranyl formate. The sample was imaged with a Tecnai T12 electron microscope equipped with a LaB$_6$ filament and operated at an acceleration voltage of 120 kV. Images were recorded using low-dose procedures at a magnification of ×52,000 and a defocus value of about −1.5 µm (Supplementary Fig. 2). Micrographs were digitized with a Zeiss SCAI scanner using a step size of 7 µm. Pixels (2 × 2 and 3 × 3) were averaged to obtain a pixel size of 2.7 Å and 4.05 Å on the specimen level for D1–D4 and D1–D8, respectively. For the projection analysis 12,093 D1–D4 particles and 5,799 D1–D8 particles were interactively selected from images using the display program WEB, which is associated with the SPIDER program suite[33]. The D1–D4 particles were windowed into 90 × 90 pixel images and the D1–D8 particles into 120 × 120 pixel images, and subjected to 10 cycles of multi-reference alignment and K-means classification specifying 50 output classes for D1–D4 and 100 classes for D1–D8.

Sequence analysis. Genomic sequence data were obtained by BLASTN and TBLASTN searches using Flyblast (http://flybase.net/blast/) and the NCBI trace archive. Genomic sequences of the following species were used for the analysis: *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba*, *D. erecta*, *D. ananassae*, *D. pseudoobscura*, *D. persimilis*, *D. willistoni*, *D. mojavensis*, *D. virilis*, *D. grimshawi*. Protein sequences were aligned by CLUSTALW[34] and sequence logos were generated with WebLogo (http://weblogo.berkeley.edu/), using a small-sample correction.

Cloning and protein expression of full-length extracellular mutant constructs for bead aggregation assays. The pIB-EC16-Fc and the pcDNA3-Dscam-1.30.30 vectors were a gift from W. Wojtowicz and L. Zipursky. The pIB-EC16-Fc contains a *Dscam* construct encoding the full-length extracellular domain with the exon combination exon 4.1, exon 6.30 and exon 9.30 (termed 1.34.30) as well as the Fc region of human IgG inserted into the pIB vector (pIB-EC16-Fc). A fragment containing Dscam immunoglobulin-domains d2–d5 was excised from pIB-EC16-Fc by restriction digest with *Mfe*I and *Aat*II. Dscam isoform 1.34.30 was amplified from a full-length Dscam construct[10], cut with *Mfe*I and *Aat*II and ligated into pIB-EC16-Fc. All point mutants for d2 and d3 were introduced by overlap PCR using Dscam 1.34.30 as a template (primer sequences are available on request). Dscam exon 6 in the pcDNA3-Dscam-1.30.30 was swapped by restriction digest cloning using *Nhe*I and *Kpn*I sites to obtain pcDNA3-Dscam-1.34.30.

S2 cells in serum-free medium (Gibco) were transfected with pIB-EC16-Fc wild-type and mutant constructs using Cellfectin (Invitrogen). After 3 days, stably transfected cells were selected for by adding 30 ug ml^{-1} Blasticidine (Sigma). After 1 week of selection, cells were expanded and the medium containing recombinant protein was harvested every two days. EC16-Fc proteins were captured and affinity-purified with protein A Sepharose (RepliGen), washed with PBS, and eluted with 100 mM glycine, pH2.8, followed by neutralization with 1 M Tris, pH8,. Protein integrity and purity was tested by Coomassie staining and protein concentrations were calculated by spectroscopy.

Hybrid isoforms were generated by overlap PCR using the isoforms 1.34.30 and 9.34.30 as template. Vectors, expression and protein purification were identical to wild-type isoforms described above. Primer and plasmid sequences are available on request.

Binding of beads to Cos cells. The binding-assay was carried out essentially as described[2] with some modifications: Cos-7 cells were grown in DMEM, 10% FCS, plated on round microscope coverslips in 24-well plates, grown to 70–80% confluency, and transfected with 300 ng pcDNA3-Dscam-1.34.30 and 130 ng pcDNA3-GFP per well using the Fugene 6 transfection reagent (Roche).

Red fluorescent sulphate polystyrene microspheres (125 µl; Duke Scientific) with a diameter of 0.39 µm and a stock concentration of 2% solids were washed in PBS and coated with a mouse anti-human IgG Fc antibody (Abazyme) overnight at 4 °C, washed again and resuspended in 250 µl of PBS, 0.1% BSA. Beads were further diluted 20-fold in PBS, 0.1% BSA and briefly sonicated. Twenty microlitres of beads were then incubated in a total volume of 100 µl PBS, 0.1% BSA with 5 µg ml^{-1} Dscam wild-type or mutant proteins for 2 h at 15 °C with 600 r.p.m. agitation. Dscam-coated beads were sonicated in a cup horn and added to the Cos cells in DMEM containing 10% FCS at a final bead

nature

concentration of 0.002% solids. Cells were allowed to bind beads for 2 h at 37 °C, followed by 5 washing steps with DMEM. Cover slips were mounted on microscope slides for imaging.

Image analysis and quantification. Cells were imaged with a Zeiss Axioskop fluorescence microscope: Random fields with GFP-positive cells were selected and corresponding images of the green (cells) and red (beads) channels were saved. Images were processed and scrambled for blind manual counting. For each field the total number of GFP-expressing cells and GFP-positive, 'binding' cells was determined. A cell was called 'binding' if it was covered with at least ten beads and if the beads resembled the cell shape (that is, a large cell binding beads in only a small area was not called 'binding'). Next, the percentage of all binding cells was calculated by adding total cells and binding cells from all fields for a given mutation. About 20% of cells were binding cells if incubated with 1.34.30 wild-type beads. To aggregate data from experiments that were performed on different days, binding was expressed as 'binding as a per cent of 1.34.30 wild-type binding'. Variation within the experiment was assessed for each mutant protein by calculating the sample standard deviation of the percentage of binding cells for each field. The total number of cells that were counted for each mutation is: K112E, H114D, 1,103; V218P, 1,408; D109A, N111A, L217A, Q219A, 1,986; N111A, K112A, K220A, P221A, 1,895; N111A, K112A, 2,313; L217A, Q219A, 1,760; E145A, N147A, M225A, V226A, 1,620; N147A, F149A, H228A, T229A; 1,755; 1.34.30 wild type, 3,938; 1.30.30 wild type, 241.

23. Otwinowski, Z. & Minor, W. in *Macromolecular Crystallography* (eds Carter, C. W. Jr & Sweet, R. M.) 307–326 (Academic Press, New York, 1997).
24. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1772–1779 (2002).
25. La Fortelle, E. & Bricogne, G. in *Methods in Enzymology*, *Macromolecular Crystallography* (eds Carter, C. W. Jr & Sweet, R. M.) 472–494 (Academic Press, New York, 1997).
26. CCP4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
27. McRee, D. E. XtalView/Xfit—A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165 (1999).
28. Cowtan, K., Zhang, K. & Main, P. in *International Tables for Crystallography* (eds Rossmann, M. & Arnold, E.) 25.2.5 (Kluwer Academic Publishers, Dordrecht, 2001).
29. Perrakis, A., Morris, R. & Lamzin, V. S. Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463 (1999).
30. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).
31. Adams, P. D. *et al.* Recent developments in the PHENIX software for automated crystallographic structure determination. *J. Synchrotron Radiat.* **11**, 53–55 (2004).
32. Ohi, M., Li, Y., Cheng, Y. & Walz, T. Negative staining and image classification—powerful tools in modern electron microscopy. *Biol. Proced. Online* **6**, 23–34 (2004).
33. Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).
34. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).