RESEARCH ARTICLE

STRUCTURAL BIOLOGY

Structural basis of Dscam1 homodimerization: Insights into context constraint for protein recognition

Shu-Ang Li,[1]* Linna Cheng,[1]* Yamei Yu,[1]† Jia-huai Wang,[2]† Qiang Chen[1]†

The *Drosophila* neural receptor Dscam1 (Down syndrome cell adhesion molecule 1) plays an essential role in neuronal wiring and self-avoidance. Dscam1 potentially encodes 19,008 ectodomains through alternative RNA splicing and exhibits exquisite isoform-specific homophilic binding, which makes it an exceptional example for studying protein binding specificity. However, structural information on Dscam1 is limited, which hinders illumination of the mechanism of Dscam1 isoform-specific recognition. Whether different Dscam1 isoforms adopt the same dimerization mode remains a subject of debate. We present 12 Dscam1 crystal structures, provide direct evidence indicating that all isoforms adopt a conserved homodimer geometry in a modular fashion, identify two mechanisms for the Ig2 binding domain to dispel electrostatic repulsion during dimerization, decode Ig2 binding specificity by a central motif at its symmetry center, uncover the role of glycosylation in Dscam1 homodimerization, and find electrostatic potential complementarity to help define the binding region and the antiparallel binding mode. We then propose a concept that the context of a protein may set restrictions to regulate its binding specificity, which provides a better understanding of protein recognition.

INTRODUCTION

Interactions between proteins are essential for most life processes, including signal transduction, protein transportation, assembly of macromolecular complexes and molecular machines, and immune response. The specificity of protein interactions is essential for the performance of protein functions and should be discriminated from nonspecific interactions.

The *Drosophila* receptor Dscam1 (Down syndrome cell adhesion molecule 1) provides us with a unique opportunity to study protein binding specificity because its ectocellular domain can potentially generate 19,008 different isoforms via alternative RNA splicing (*1*) and exhibits exquisite isoform-specific binding to form homodimers (*2*). Extensive studies have demonstrated that Dscam1 plays important roles in neural wiring and self-avoidance (*3*, *4*) and that isoform-specific homophilic recognition is the molecular basis of its functions. Accumulating evidence has shown that diversity is essential for Dscam1's functions (*5*–*8*). Mimicking the role of Dscam1 in neuronal self-recognition and self-avoidance, the mammalian clustered protocadherins also require isoform diversity (*9*). Specificity plays a particularly central role for such molecules to perform their functions, and maintaining the specificity of the vast number of isoforms is a great challenge.

Dscam1 contains 10 immunoglobulin (Ig) domains, six fibronectin type III repeats, a single transmembrane segment, and a C-terminal cytoplasmic domain. Homophilic binding requires three variable Ig domains (Ig2, Ig3, and Ig7) and direct binding of the matching Ig

domains (Ig2/Ig2, Ig3/Ig3, and Ig7/Ig7) (Fig. 1, A and B). Dscam1 Ig2, Ig3, and Ig7 have 12, 48, and 33 different isoforms, respectively. Different isoforms share a 46 to 81%, 40 to 88%, and 32 to 93% identity of protein sequence for the variable part of Ig2, Ig3, and Ig7, respectively. To date, only two isoforms, three isoforms, and one isoform of Ig2, Ig3, and Ig7, respectively, have published structures. The lack of structural comparison between different isoforms hinders researchers from elucidating the detailed mechanism of Dscam1 binding specificity. The structure of Dscam1 Ig1-Ig8 domains reveals that it is an S-shaped molecule and forms a symmetric homodimer, suggesting that all isoforms adopt the same dimerization mode (*10*). However, a previous study showed that Ig1-Ig4 isoforms could adopt different binding modes for homodimerization (*11*). Although biochemical data support the idea that the three variable Ig domains interact in an independently modular fashion (*2*), direct structural evidence is still lacking. More Dscam1 structures are needed to clarify these discrepancies and ambiguities.

Here, we present 10 crystal structures of *Drosophila* Dscam1 Ig1-Ig4 (1.90 to 4.00 Å) and 2 crystal structures of Ig7 (1.95 and 2.37 Å). We carried out an extensive comparison of these structures and revealed the detailed mechanism of Dscam1 homobinding specificity. Two mechanisms were adopted by Ig2 to dispel electrostatic repulsion during dimerization. A central motif at the Ig2 binding site mainly determined the binding specificity of Ig2. The Ig domain consists of a two-layer sandwich of antiparallel β strands, ABED and CFG, respectively. A special electrostatic potential pattern on the ABED face of Ig7 helped ensure the antiparallel binding of the Ig7 homodimer. Glycosylation regulated the formation of the Dscam1 homodimer. We provided direct evidence showing that different Dscam1 isoforms adopt the same dimerization mode in a modular fashion and that the different binding mode of Ig1-Ig4$_{9,9}$ (*11*) is a crystallization artifact. Similar nonphysiological interactions occurred in the crystal structures of phosphodiesterase 2 (*12*, *13*), Vps75 (*14*, *15*), Ca^{2+}/calmodulin-dependent protein

[1]State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, Sichuan University, and Collaborative Innovation Center of Biotherapy, Chengdu 610041, P. R. China. [2]Departments of Medical Oncology and Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA. Department of Pediatrics and Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: qiang_chen@scu.edu.cn (Q.C.); jwang@crystal.dfci.harvard.edu (J.-h.W.); yamei_yu@scu.edu.cn (Y.Y.)

kinase II (CaMKII) (16, 17), and p75 (18, 19). Therefore, we propose the concept that a multidomain protein may set restrictions to each single domain so that, within the context, only some specific binding modes are available for interaction. Glycosylation may also have a similar function to regulate protein binding. Such restrictions, namely, context constraints, help avoid nonspecific interactions for protein recognition.

RESULTS

Structures of various Dscam1 isoforms are determined for comparison

Dscam1 Ig2 has 12 isoforms, and the structures of 6 of these isoforms (isoforms 1, 4, 6, 7, 8, and 9) were solved in this study. To facilitate comparison, we made Dscam1 Ig1-Ig4 constructs using a certain Ig2 isoform combined with various Ig3 isoforms, or vice versa. Eight Ig1-Ig4 structures (isoforms 1.9, 4.4, 4.44, 6.9, 6.44, 7.44, 8.4, and 9.44; the first number indicates the Ig2 isoform, and the second number indicates the Ig3 isoform) and two Ig7 structures (isoforms 5 and 9) were solved and refined. The Ig1-Ig4$_{1.34}$ [Protein Data Bank (PDB) ID 2V5M] structure was re-refined to complete the glycan model. Details of the structure refinement statistics are summarized in table S1.

Different Dscam1 isoforms share the same dimerization mode in a modular fashion

The three variable Ig domains have been proposed to interact independently in a modular fashion (2). Modular interactions of the variable domains provide a molecular strategy for achieving remarkable homophilic specificity in Dscam1. However, direct structural evidence is still lacking. Here, we made Dscam1 Ig1-Ig4 constructs using a certain Ig2 isoform combined with various Ig3 isoforms, or vice versa. Comparison of different Ig1-Ig4 structures showed that the Ig2 dimer interface remained the same when combined with different Ig3 isoforms and that the Ig3 dimer interface also remained the same when combined with different Ig2 isoforms (Fig. 1C). Thus, binding at each pair of identical variable Ig domains occurs independently.

Whether different Dscam1 isoforms adopt the same dimerization mode is a subject of debate (10, 11). The Ig1-Ig4$_{9.9}$ dimer exhibits a binding mode different from that observed in Dscam1 structures Ig1-Ig4$_{1.34}$ and Ig1-Ig8$_{1.30.30}$ (10, 11). To investigate this discrepancy, we combined different Ig2 isoforms (Ig2$_1$ and Ig2$_6$) with Ig3$_9$, and solved the structures of Ig1-Ig4$_{1.9}$ and Ig1-Ig4$_{6.9}$. We also combined a different Ig3 (Ig3$_{44}$) with Ig2$_9$, and solved the Ig1-Ig4$_{9.44}$ structure. These structures, together with the structures of the other five Ig1-Ig4 isoforms presented here, could be well superpositioned, and all adopted the same dimerization mode
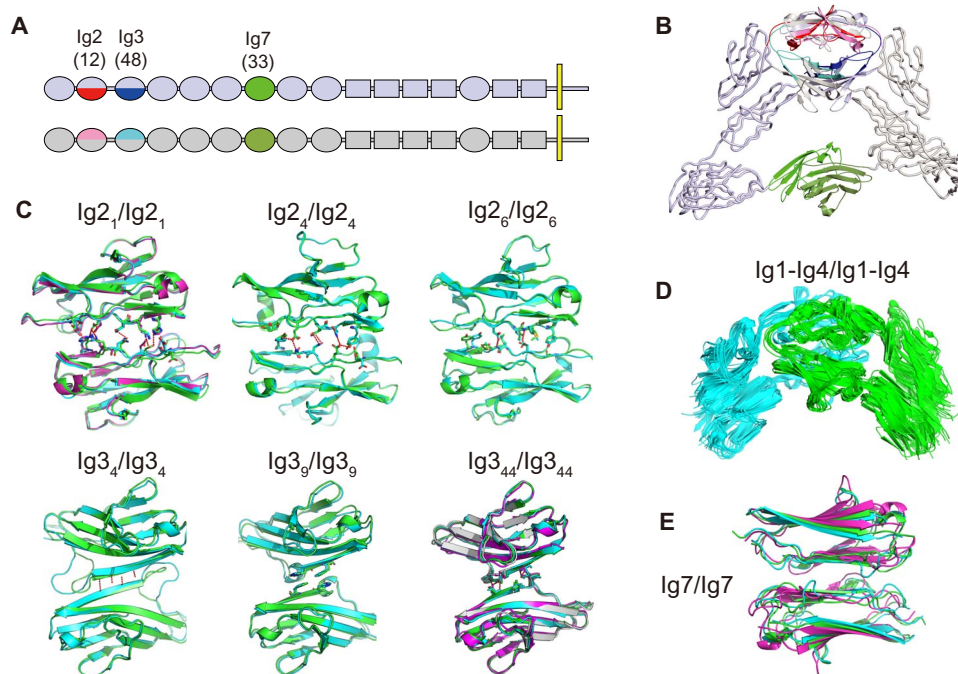

Fig. 1. Dscam1 isoforms adopt the homodimer geometry in a modular fashion. (**A**) Dscam1 as a cell surface receptor comprising 10 Ig domains (ovals), 6 fibronectin type III domains (rectangles), a transmembrane domain (yellow line), and a cytoplasmic tail. Dscam1 proteins engage in isoform-specific homophilic binding, which is determined by the variable parts encoding the N-terminal halves of Ig2 (red or pink; 12 isoforms) and Ig3 (blue or cyan; 48 isoforms), and all of Ig7 (light green or dark green; 33 isoforms). (**B**) Dscam1 homodimer structure (PDB ID 3DMK). The domains are colored as in (A). The homophilic interaction of Dscam1 is mediated by the direct binding of the matching variable Ig domains (Ig2/Ig2, Ig3/Ig3, and Ig7/Ig7). (**C**) Superpositions of Dscam1 Ig2 dimers or Ig3 dimers, viewed from the top of (B). The same Ig2 dimers in combination with different Ig3 are superimposed: isoform 1(green for 1.9, cyan for 1.30, and purple for 1.34), isoform 4 (green for 4.4 and cyan for 4.44), and isoform 6 (green for 6.9 and cyan for 6.44). The same Ig3 dimers in combination with different Ig2 are superimposed: isoform 4 (green for 4.4 and cyan for 8.4), isoform 9 (green for 1.9 and cyan for 6.9), and isoform 44 (green for 4.44, cyan for 6.44, silver for 7.44, and purple for 9.44). (**D**) Superposition of Dscam1 Ig1-Ig4 dimers, viewed from the same orientation of (B). Ten different Ig1-Ig4 isoforms (1.9, 1.30, 1.34, 4.4, 4.44, 6.9, 6.44, 7.44, 8.4, and 9.44/Zn) are superimposed on the basis of the Ig2-Ig3 dimer of molecule A (cyan). The chains of molecule B (green) overlap closely. (**E**) Superposition of Dscam1 Ig7 dimers, viewed from the bottom of (B). Three Ig7 dimers are superimposed (green, cyan, and purple for isoforms 5, 9, and 30, respectively).

as that in Ig1-Ig4$_{1.34}$ and Ig1-Ig8$_{1.30.30}$ (Fig. 1D). We carried out a pairwise comparison of different isoforms by superpositioning Ig2-Ig3 of molecule A and calculated the intermolecular angle of molecule B. The small intermolecular angles (0.4° to 10.9°) indicated that they shared the same binding mode (table S2). Different Ig7 isoforms also shared the same homodimerization mode (Fig. 1E). The Ig2$_9$/Ig2$_9$ or Ig3$_9$/Ig3$_9$ interface presented here is different from that observed in the Ig1-Ig4$_{9.9}$ dimer (11) (Fig. 1C). Our results support the concept that homodimer geometry is conserved in all Dscam1 isoforms and that the Ig1-Ig4$_{9.9}$ dimer is a crystallization artifact.

Collectively, comparison of the structures of different Dscam1 isoforms provided direct evidence confirming that different isoforms shared the same dimerization mode and that the three variable Ig domains recognized themselves via a modular fashion.

Ig2/Ig2 interface: Two mechanisms to dispel electrostatic repulsion

Dscam1 shows exquisite isoform-specific binding to form homodimers (2). Specificity has been proposed to be dependent on both electrostatic and shape complementarity, and homophilic interface docking models have been made for each of the 12 Ig2 isoforms (10).

Two Ig2 isoforms, Ig2$_8$ and Ig2$_9$, have an acidic residue (E or D) at their symmetric centers, which are supposed to result in electrostatic repulsion and to disrupt homodimer formation. In the docking models, the side chains of the symmetrically central residue E or D were swung away as far as possible (10). We solved the crystal structures of Ig1-Ig4$_{8.4}$ and Ig1-Ig4$_{9.44}$ (with or without bivalent cation zinc) and found that Ig2$_8$ and Ig2$_9$ used two different mechanisms to dispel electrostatic repulsion during dimerization.

In the Ig2$_8$ dimer, K126 on the BC loop protruded to the symmetric center and formed a salt bridge with the symmetric center E111 to neutralize the negative charge of E111 (Fig. 2A). In the Ig2$_9$/Zn dimer, E126 (the residue corresponding to K126 in Ig2$_8$), together with E109 and D111, coordinated with two zinc ions via their bidentate carboxylate groups (Fig. 2A). Either via formation of a salt bridge or via coordination with bivalent cations, the negative charge on the symmetric center residue was dispelled, and Ig2 homodimers of isoforms 8 and 9 were allowed to form in the same mode as other Ig2 isoforms. Although
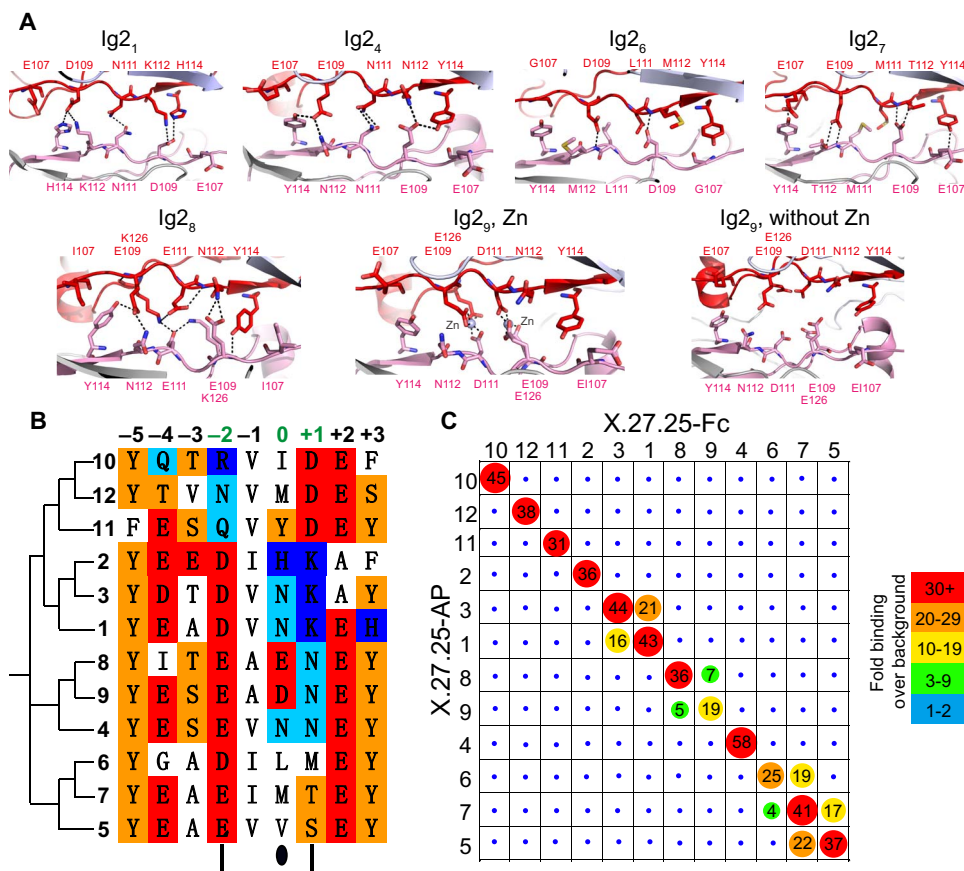

Fig. 2. A central motif mainly determines Dscam1 Ig2 specificity. (A) Details of Dscam1 Ig2 homodimer interfaces. Color coding as in Fig. 1. The residues involved in the homophilic interactions between Ig2 dimers are shown as sticks and labeled. Hydrogen bonds are drawn as black dashed lines. **(B)** Sequence alignment of Ig2 interface residues. Color-coded for hydrophobic groups (white), acidic groups (red), basic groups (blue), polar hydroxyl groups (orange), and polar amino groups (cyan). Black ellipse marks the symmetric center (residue position 0). The residue at position −2 always interacts with the residue at position +1. This motif consists of three residues at positions −2, 0, and +1 (green), and mainly determines the specificity of Dscam1 Ig2. **(C)** Binding properties of all Dscam1 Ig2 variable domains [data were adapted from Wojtowicz et al. (2)]. Binding is indicated as fold over background by the number and a color scale. The sizes of the balls are scaled to the numbers in each block.

Ig1-Ig4$_{9.44}$ without bivalent cations formed a similar dimer as the Ig1-Ig4$_{9.44}$-Zn structure (fig. S1), the Ig2$_9$ dimer did not form properly as a result of the negative electrostatic repulsion at its symmetric center (Fig. 2A). Surface plasmon resonance (SPR) measurement showed that the homophilic binding of Ig1-Ig4$_{9.44}$ was significantly enhanced in the presence of zinc acetate (fig. S2).

Although very similar to the docking models (10), the Ig2 dimer interfaces for isoforms 4 and 6 showed some differences. In the Ig2$_4$ docking model, N111 formed a hydrogen bond with E109. However, in the crystal structure of the Ig2$_4$ dimer, N111 formed two hydrogen bonds with N111 from the other monomer, and E109 formed hydrogen bonds with N112 and Y114 from the other monomer (Fig. 2A). For the Ig2$_6$ dimer, D109 could not form the hydrogen bond with Y114 as in the docking model; instead, it formed a hydrogen bond with the M112 main chain in the crystal structure (Fig. 2A).

Decoding the binding specificity of Dscam1 Ig2

The Ig2/Ig2 interface is twofold symmetric, mainly composed of the A-A′ segment that is oriented in an antiparallel fashion. The nomenclature we used for the Ig2 interacting segment designated the symmetrically central residue as position 0. The residues flanking this site were designated as positions +1, −1, and so forth, as indicated in Fig. 2B. Ig domains share a common core Greek-key β-sandwich structure. Along the β strand, the neighboring residues' side chains usually adopt two alternative orientations: inward and outward. In all the known structures of Dscam1 Ig2, the consecutive residues at positions 0 and +1 have an outward-oriented side chain, making a bulge, and both are involved in the Ig2/Ig2 dimerization contacts. An essential finding is that the residue at position +1 always interacted with the residue at position −2 (Fig. 2A) (isoform 9 is an exception because its E109 at position −2 participated in the cationic coordination). We found that the three residues at positions −2, 0, and +1 could mainly determine the specificity of all the Ig2 isoforms.

Isoforms 10, 11, and 12 have a basic or polar amino residue at position −2 and, complementarily, an acidic or polar hydroxyl residue at position +1. Vice versa, isoforms 1 to 4, 8, and 9 have an acidic or polar hydroxyl residue at position −2 and, complementarily, a basic or polar amino residue at position +1 (Fig. 2B). If the residue D at position −2 (owning a shorter side chain; isoforms 1 to 3) is replaced by the residue E (owning a longer side chain; isoform 4, 8, or 9), correspondingly, the residue K at position +1 (owning a longer side chain; isoforms 1 to 3) is replaced by the residue N (owning a shorter side chain; isoform 4, 8, or 9) (Fig. 2B). The complementarity of both electrostaticity and length confers on this central motif the ability to distinguish between different isoforms.

As to the issue of cross-interaction, isoforms 1 and 3 have an identical central motif (D−2, N0, and K+1) at the symmetric center; therefore, they have a strong cross-interaction (Fig. 2, B and C). Although having the same residue at positions −2 and +1 as isoforms 1 and 3, isoform 2 has a different residue at position 0 and made no cross-interaction with isoform 1 or 3. Isoforms 8 and 9 have a very similar central motif; however, they adopted two different mechanisms to solve the electrostatic repulsion problem and only showed minor cross-interaction. The cross-interaction between isoforms 5 and 7 is more significant than that between isoforms 6 and 7, which is readily explained by the fact that the central motif is more similar between isoforms 5 and 7 than between isoforms 6 and 7 (Fig. 2, B and C).

In summary, on the basis of the structural analysis, we proposed that a central motif at the symmetric center of Dscam1 Ig2 homodimer interfaces mainly determined the specificity of Ig2. This was supported by recently published in vivo data, which indicate that altering the charge of residues at Ig2 position −2 or +1 converted their binding specificity (20).

Ig3/Ig3 interface: Smaller but more variable

The Ig3/Ig3 interface was twofold symmetric, composed of the A-A′ segment in an antiparallel fashion. In all known Dscam1 structures, the Ig3/Ig3 interface contributed a much less buried surface area than the Ig2/Ig2 interface (table S3).

To confirm that the Ig3$_9$ homodimer in the Ig1-Ig4$_{9.9}$ structure is an artifact, we solved the structures of Ig3$_9$ combined with two other Ig2 isoforms (isoforms 1 and 6). The physiological Ig3$_9$ dimer interface remained the same in these two structures and was composed of two pairs of His/Phe π-π interactions (Fig. 3). The physiological Ig3$_9$ homodimer has a much larger buried surface area compared with the artificial Ig3$_9$ homodimer (table S3).

Ig3$_4$ is the longest isoform for Ig3 and buried the largest surface area among known Ig3 structures (table S3). From H225 to L228 (number as in Ig1-Ig4$_{8.4}$), the main chains of the two protomers paired into an antiparallel β sheet, which continued to the A′ strand and made an extra long A′ strand (Fig. 3). The A strand of Ig3$_4$ formed a long loop and adopted different conformations in the structures of Ig1-Ig4$_{4.4}$ and Ig1-Ig4$_{8.4}$ (fig. S3A). This may reflect the flexibility of the Ig3$_4$ A strand.

The interface of the Ig3$_{44}$ dimer consisted of the central Ile^{220} (number as in Ig1-Ig4$_{7.44}$) hydrophobic interaction and two pairs of intercellular hydrogen bonds (Fig. 3).

The A-A′ segment in Ig3 is more flexible and can adopt different conformations compared to that in Ig2. Burying a smaller surface area compared to the Ig2/Ig2 homodimer, the Ig3/Ig3 dimers seem more variable than the Ig2/Ig2 dimers. A common feature we have observed is that the available Ig3 dimers all involved hydrophobic interactions at their interfaces (Fig. 3) (10, 11).
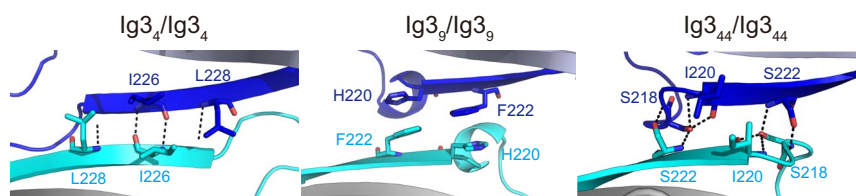

Fig. 3. Dscam1 Ig3 homodimer interfaces show more variability. The residues involved in the homophilic interactions between Ig3 dimers are shown as sticks and labeled. Color coding as in Fig. 1. Hydrogen bonds are drawn as black dashed lines.

Li *et al. Sci. Adv.* 2016;2:e1501118   27 May 2016

4 of 9

Ig7/Ig7 interface: Surface features ensure antiparallel binding

Among the three variable Ig domains, Ig7 has shown much more variability than Ig2 or Ig3. First, the whole Ig7 domain is variable, whereas only half of Ig2 or Ig3 is variable. Second, the expression of Ig7 shows more changes both temporally and spatially during development (21). Hitherto, only one among 33 Ig7 isoforms has structure information available (isoform 30) (10). Here, we present the crystal structures of the other two Ig7 isoforms (isoforms 5 and 9).

Similar to the structure of $Ig7_{30}$, $Ig7_5$ and $Ig7_9$ also use the ABED face to form an antiparallel homodimer. The most interesting finding is that all the known Ig7 structures have a complementary electrostatic potential surface pattern on the ABED face: positive in one end, neutral in the middle, and negative in the other end (Fig. 4A). This pattern en-sures that Ig7 dimerizes with a specific antiparallel mode. We have observed conserved acidic residues on the AB loop and the EF loop, and basic residues on the BC loop and the DE loop (Fig. 4B). Sequence alignment indicates that this complementary electrostatic potential surface pattern is presented in most, if not all, Ig7 isoforms (Fig. 4C).

The role of glycosylation in Dscam1 dimerization

Dscam1 Ig1-Ig4 has two constant N-glycosylation sites: one is on the Ig1 A′B loop and the other is on the Ig3 F strand (Fig. 5A). We found that the glycan on Ig3 interacted directly with the Ig4 domain of the other monomer. In the Ig1-Ig4$_{6.9}$ structure, D95 formed salt bridges with R374 and well-positioned the N-epsilon of R374 to form hydrogen bonds with the glycan (Fig. 5A). K338 also formed hydrogen bonds with the glycan. These interactions between glycan and protein
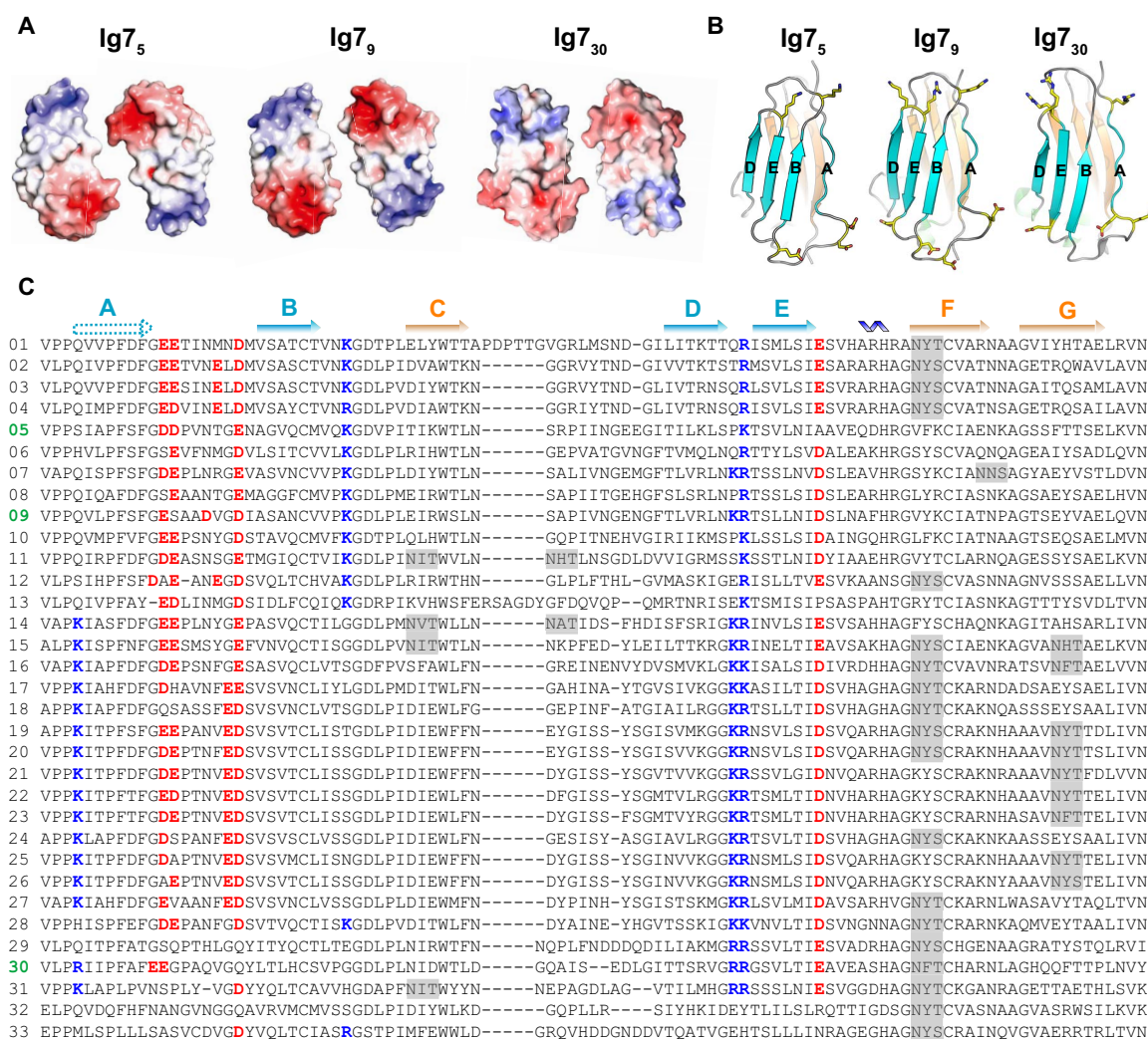

Fig. 4. Electrostatic surface potential and glycosylation help ensure the antiparallel binding of the Dscam1 Ig7 homodimer. (**A**) Open-book view of the electrostatic surface potential of the homodimers of Ig7 isoforms 5, 9, and 30. Blue and red (±5 kT/e) indicate the positively and negatively charged areas of the protein, respectively. (**B**) The ABED face of Ig7 isoforms 5, 9, and 30. The ABED face and the CFG face are presented in cyan and orange, respectively. The charged residues located at both ends of this face are shown as sticks. (**C**) Sequence alignment of all 33 Dscam1 Ig7 isoforms. The secondary structural elements are marked in accordance with the structure of Dscam1 Ig7$_5$. The positively and negatively charged residues at the ends of the ABED face are presented in bold blue and bold red, respectively. The potential N-linked glycosylation motifs are shaded in gray.

helped prevent the two molecules from sliding across a relatively flat interface. Such interactions have not been reported in previous Dscam1 structures because of low resolution or improper model building. The Ig1-Ig4$_{1.34}$ structure has a resolution of 1.95 Å; however, several water molecules have been modeled onto the glycan positions (11). We performed re-refinement of the Ig1-Ig4$_{1.34}$ structure (PDB ID 2V5M) to complete the glycan models, and we found that it also interacted with the other monomer in exactly the same fashion observed in the Ig1-Ig4$_{6.9}$ structure (Fig. 5A). The glycan that we modeled into the Ig1-Ig4$_{1.34}$ structure had clear electronic density and could be well superpositioned with the glycan in the Ig1-Ig4$_{6.9}$ structure (Fig. 5, B and C). The two glycosylation sites on Ig1-Ig4 and the involved residues are all on the constant part and thus should be conserved among all Dscam1 isoforms. The re-refinement of the Ig1-Ig4$_{1.34}$ structure confirmed the conservation of the interactions between the glycan and the Ig4 domain.
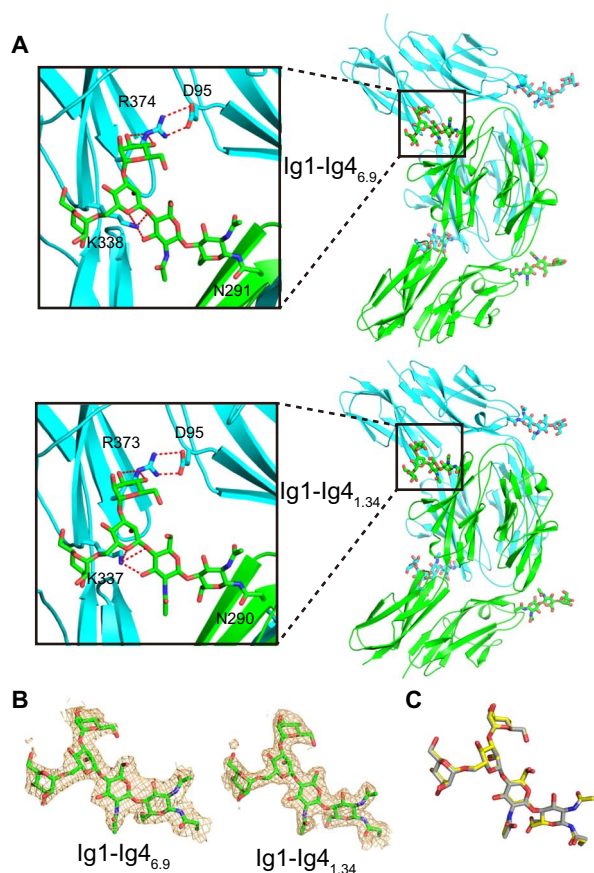

Fig. 5. Glycosylation regulates Dscam1 dimerization. (A) Structures of Dscam1 Ig1-Ig4$_{6.9}$ and Ig1-Ig4$_{1.34}$ homodimers. Structures are shown in cartoons. Glycans are shown in sticks. Two monomers are presented in cyan and green, respectively. The glycan on an invariable Asn residue of Ig3 directly interacts with the other monomer, and the details are shown in the enlarged part. The hydrogen bonds were drawn as red dashed lines. **(B)** Electron density for the glycans located on Ig3 in the Ig1-Ig4$_{6.9}$ and Ig1-Ig4$_{1.34}$ structures. The $2F_o - F_c$ electron density maps are contoured at 1σ. **(C)** Superposition of the glycans located on Ig3 in the Ig1-Ig4$_{6.9}$ and Ig1-Ig4$_{1.34}$ structures. The glycans are shown as sticks. Carbon atoms of glycans in Ig1-Ig4$_{6.9}$ and Ig1-Ig4$_{1.34}$ are presented in yellow and gray, respectively.

As to the variable part, two isoforms of Ig2 (isoforms 8 and 11) and three isoforms of Ig3 (isoforms 19, 20, and 21) have potential glycosylation sites, and they are located on the CD loop or the A′B loop (fig. S4), outside the dimer interface. Most Ig7 isoforms have potential glycosylation sites; without exception, they are located at the CFG face, the reverse side of the homophilic binding interface (Fig. 4C). Glycosylation on the back prevents the reverse side from involving homophilic binding and helps define the interface of the Dscam1 homodimer.

Context constraint for protein recognition

The artificial Ig1-Ig4$_{9.9}$ dimer structure inspired us to consider why it made such nonphysiological interactions. Superimposition of two Ig1-Ig7$_{1.30.30}$ molecules onto the Ig1-Ig4$_{9.9}$ dimer showed overlapping of Ig6 domains and distancing of Ig7 domains (fig. S5A). Apparently, Dscam1 molecules arranged as an Ig1-Ig4$_{9.9}$ homodimer could not allow the proper formation of the Ig7/Ig7 interaction.

A segment such as Ig1-Ig4$_{9.9}$ (contains two variable Ig domains) adopting a nonspecific binding mode implies that all three variable Ig domains are required to ensure a physiological interaction for Dscam1. If Ig7 dimerizes properly, the binding mode of Ig1-Ig4$_{9.9}$ will be excluded. In other words, the matching of the three variable Ig domains set restrictions for each other to exclude nonspecific binding. It is the context that ensures the correct protein recognition. Therefore, we introduced a concept—context constraint—to emphasize that the integrity of a protein may play an important role in the regulation of its specificity. More details are discussed in the following section.

DISCUSSION

Drosophila Dscam1 can generate (via RNA splicing) a vast repertoire of isoforms, which exhibit striking isoform-specific homophilic binding. The specificity of Dscam1 is determined by three variable Ig domains, namely, Ig2, Ig3, and Ig7. Maintaining specificity is essential to Dscam1 functions and is also a great challenge for such a great number of isoforms. We present 12 crystal structures and carry out an extensive comparison focusing on the three variable Ig domains, revealing the detailed mechanism of Dscam1 binding specificity. Our results show that the previously reported Ig1-Ig4$_{9.9}$ structure is an artifact forming a nonspecific dimer, which inspires us to propose a concept—context constraint—to emphasize that protein integrity may exclude nonspecific binding in protein-protein recognition.

Insights into the specificity of the Dscam1 homodimer

Biochemical studies have indicated that the Dscam1 variable Ig domain performs self-binding in a modular fashion (2). The structures we present here make it possible for us to perform an extensive structural comparison of different isoforms and to confirm that different isoforms adopt the same dimerization mode. Direct structural evidence from the comparison indicates that the variable Ig2 and Ig3 domains function independently of each other, because combination with different isoforms has not affected their dimer interfaces—especially the Ig3$_{44}$ dimer interface, which remains the same in combination with four different Ig2 isoforms (Fig. 1C). The buried surface area of the same Ig2 or Ig3 shows some variability in different combinations. The Ig2$_1$ dimer buries a smaller surface area in Ig1-Ig4$_{1.9}$ compared with that in Ig1-Ig4$_{1.34}$ and

Ig1-Ig4$_{1.30}$, and such situation occurs for Ig3$_6$ in Ig1-Ig4$_{6.9}$ compared with Ig1-Ig4$_{1.9}$ (table S3). This is attributable to the different side-chain conformations adopted by the residues around the dimer interface (fig. S3, B and C).

Our results suggest that Ig2's specificity is mainly determined by a central motif at the homodimer symmetric center (Fig. 2). A recent in vivo study supports this by showing that chimeric isoforms altering the charge of residues at Ig2 position −2 or +1 destroy their homophilic binding and that complementary chimeras restore binding and rescue Dscam1 function (20).

Our results reveal that the Dscam1 homodimer has conspicuous complementarity. A central motif at the symmetric center mainly determines the specificity of Ig2. Two residues in this motif (positions −2 and +1) have side chains characteristic of complementary electrostaticity and length (Fig. 2B). Ig7 shows a complementary electrostatic potential on its homophilic binding face (Fig. 4). More studies are needed to clarify whether isoforms 32 and 33 of Ig7 have similar electrostatic potential patterns.

Ig1-Ig4$_{9.9}$ homodimer is an artifact

We demonstrate that Ig2$_9$ needs bivalent cations to neutralize the negative charge at the homodimer symmetric center and adopts the same dimerization mode as other Ig2 dimers, except that in Ig1-Ig4$_{9.9}$. In the structure of Ig1-Ig4$_{9.44}$ without bivalent cations, Ig2$_9$ was repulsed apart (Fig. 2A). The crystals of Ig1-Ig4$_{9.44}$ without bivalent cations could only be obtained at low pH (pH 4.6) because low pH could reduce the negative charge and, to some extent, depress the electrostatic repulsion. The Ig3$_9$ dimer remains the same in combination with two different Ig2 isoforms (isoforms 1 and 6) and differs from that in the Ig1-Ig4$_{9.9}$ structure (Figs. 1C and 3). Superpositioning of the Ig1-Ig8 structure on the Ig1-Ig4$_{9.9}$ dimer shows that the arrangement of the Ig1-Ig4$_{9.9}$ homodimer is incompatible with the proper Ig7/Ig7 interaction (fig. S5A). These results provide direct evidence indicating that the Ig1-Ig4$_{9.9}$ structure is artificial as a result of a lack of bivalent cations.

Glycosylation regulates protein binding

The Dscam1 Ig1-Ig8 structure has shown that Ig1 contacts with the loop connecting Ig2 and Ig3, which acts as a "bookend" to stop the molecules from sliding over the proper position (10). Our structures show that the two glycosylation sites on Ig1 and Ig3 locate on either side of the bookend, and that especially the glycan on Ig3 directly interacts with Ig4 of the other monomer and plays a similar role to help define the binding region (Fig. 5).

All the potential glycosylation sites on the variable Ig domains are on the back or side of the binding face and thus help restrict homophilic binding to the interface we have observed in Dscam1 homodimer structures. Similar situations could be found in other proteins such as intercellular adhesion molecule–3 (22) and Del-1 (23).

Glycosylation has been found to affect protein recognition for the common neurotrophin receptor p75. The ectodomain of glycosylated p75 formed a symmetric 2:2 ligand/receptor complex (18), whereas the ectodomain of nonglycosylated p75 formed an asymmetric 2:1 ligand/receptor complex (19) (fig. S5E). The 2:1 complex has been shown to be the result of artificial nonglycosylation (18). Glycans serve a variety of structural and functional roles (24), and N-linked glycosylation of p75 has been shown to play an important role in nerve growth factor signaling (25). The differences observed in the ligand binding between glycosylated and nonglycosylated p75 underline the importance of protein integrity in protein-protein interactions.

Context constraint

Specificity involves not only binding to a specific partner but also not binding to other proteins, namely, it favors a small set of interactions over a multitude of possibilities. In the cell, proteins compose a significant proportion of dry mass: 55% in Escherichia coli (26) and 39.6% in Saccharomyces cerevisiae (27). A protein generally resides in a crowded environment, and maintaining specific protein interactions is an indispensable requirement for cell function.

Most genomic proteins (about 65% in archaea or bacteria and 80% in eukaryotes) are multidomain proteins created through gene duplication, recombination, and fusion (28). In a multidomain protein, each domain may fulfill its own function either independently or in a concerted manner with its neighbors. Here, we propose that a multidomain protein may set a context for each single domain, so that the binding modes for protein-protein interactions are restricted to some specific ones. Such restrictions may play important roles in avoiding nonspecific interactions.

Eight Dscam1 N-terminal Ig domains adopt an S-shaped configuration and form a symmetric homodimer, and any Dscam1 variant was compatible with the double "S" scaffold of this Dscam1 Ig1-Ig8 homophilic dimer (10). Ig1-Ig4$_{9.9}$ forms an artificial dimerization arrangement because of a lack of bivalent cations and is not compatible with the double "S" scaffold (fig. S5A). If Ig7 is involved in the dimerization, nonspecific binding of Ig1-Ig4$_{9.9}$ should be avoided. The structure of Dscam1 Ig1-Ig8$_{1.30.30}$ also shows flexibility in the hinges between domains, which results in a 13-Å shift between Ig7 and the Ig1-Ig4 horseshoe (10). Whether this reflects the intrinsic flexibility of Dscam1 or the lack of the context constraint of the whole ectocellular domain needs further studies.

Similar situations may be found in other proteins such as phosphodiesterase 2A (12, 13), Rtt109 (regulator of Ty1 transposition 109) (14, 15), and CaMKII (16, 17). These examples clearly demonstrate the restrictions set by a full-length context to the binding mode of its truncated part (fig. S5, B to D).

Our results and those of the previous reports mentioned above reinforce the importance of protein integrity in protein recognition. We propose that the context of a protein sets restrictions for its recognition. When nature brings several domains together to form multidomain and multifunctional proteins, or does posttranslational modifications to proteins, it may endow them with intrinsic restrictions that regulate the specificity of protein interactions. The context constraint is required, at least for some proteins, to ensure the specificity of physiological binding.

MATERIALS AND METHODS

Protein production and crystallization

Protein production and crystallization were performed, as previously described (29, 30). Briefly, all the Ig1-Ig4 constructs were expressed in a baculovirus system (Bac-N-Blue; Invitrogen), and the two Ig7 constructs were expressed in E. coli strain Rossetta (Novagen) or SHuffle (NEB). Purified Dscam1 protein samples are shown in fig. S6. Crystals were grown at 289 K using the hanging-drop vapor diffusion method.

Data collection and processing

Diffraction data were collected on beamline 3W1A at the Beijing Synchrotron Radiation Facility (BSRF) and on beamline BL17U1 at the Shanghai Synchrotron Radiation Facility (SSRF). Collected data were processed by X-ray Detector Software or HKL2000. Details had been previously described (29, 30).

Structure determination and refinement

The structures were determined by molecular replacement using the Ig1-Ig4 (PDB ID 2V5M) or the Ig7 segment of the Dscam1 Ig1-Ig8 structure (PDB ID 3DMK) as search model. Structure refinement and model building were performed with PHENIX (31) and Coot (32). All models were validated with MolProbity (33). The refinement results and PDB IDs are summarized in table S1. All structure figures were prepared with PyMOL (www.pymol.org). Buried surface area was calculated at the PISA Web server (www.ebi.ac.uk/msd-srv/prot_int/pistart.html).

Intermolecular angle calculation

Every Dscam1 Ig2-Ig3 pair being compared was superpositioned by the SUPER command in PyMOL. The matrix generated by a second superposition on the second tandem domain was obtained with the GET_OBJECT_MATRIX command in PyMOL and used to calculate the κ angle in polar angles.

SPR analysis

The binding of Dscam1 Ig1-Ig4$_{9.44}$ was assessed by SPR on the Biacore X100 (GE Healthcare) at room temperature. Purified Ig1-Ig4$_{9.44}$ protein was covalently immobilized to the surface of a CM5 sensor chip using the Amine Coupling Kit Biacore (GE Healthcare). Purified Ig1-Ig4$_{9.44}$ protein in the absence or presence of 0.2 mM zinc acetate was flowed over the sensor chip in running buffer [20 mM tris-HCl (pH 7.4), 100 mM NaCl, and 0.5% (v/v) Tween 20] at a flow rate of 10 μl/min. The chip surface was regenerated by 10 mM glycine-HCl (pH 1.7).

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/2/5/e1501118/DC1

fig. S1. Same dimerization mode in Dscam1 Ig1-Ig4$_{9.44}$ with or without bivalent cations.
fig. S2. SPR measurement of Ig1-Ig4$_{9.44}$ homophilic binding.
fig. S3. Variability of the same Ig2 or Ig3 in different combinations.
fig. S4. Potential glycosylation sites on the variable parts of Dscam1 Ig2 and Ig3.
fig. S5. Context setting restrictions for protein recognition.
fig. S6. Purified Dscam1 proteins for crystallization.
table S1. X-ray data refinement statistics.
table S2. Intermolecular angles (°) of different Dscam1 Ig2-Ig3 isoforms.
table S3. Buried area (Å^2) of the Dscam1 homodimer of Ig1-Ig4, Ig2, and Ig3.

REFERENCES AND NOTES

1. D. Schmucker, J. C. Clemens, H. Shu, C. A. Worby, J. Xiao, M. Muda, J. E. Dixon, S. L. Zipursky, *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* **101**, 671–684 (2000).
2. W. M. Wojtowicz, W. Wu, I. Andre, B. Qian, D. Baker, S. L. Zipursky, A vast repertoire of Dscam binding specificities arises from modular interactions of variable Ig domains. *Cell* **130**, 1134–1145 (2007).
3. S. L. Zipursky, W. B. Grueber, The molecular basis of self-avoidance. *Annu. Rev. Neurosci.* **36**, 547–568 (2013).
4. S. S. Millard, S. L. Zipursky, Dscam-mediated repulsion controls tiling and self-avoidance. *Curr. Opin. Neurobiol.* **18**, 84–89 (2008).
5. D. Hattori, E. Demir, H. W. Kim, E. Viragh, S. L. Zipursky, B. J. Dickson, Dscam diversity is essential for neuronal wiring and self-recognition. *Nature* **449**, 223–227 (2007).
6. D. Hattori, Y. Chen, B. J. Matthews, L. Salwinski, C. Sabatti, W. B. Grueber, S. L. Zipursky, Robust discrimination between self and non-self neurites requires thousands of Dscam1 isoforms. *Nature* **461**, 644–648 (2009).
7. H. He, Y. Kise, A. Izadifar, O. Urwyler, D. Ayaz, A. Parthasarthy, B. Yan, M.-L. Erfurth, D. Dascenco, D. Schmucker, Cell-intrinsic requirement of Dscam1 isoform diversity for axon collateral formation. *Science* **344**, 1182–1186 (2014).
8. B. E. Chen, M. Kondo, A. Garnier, F. L. Watson, R. Puettmann-Holgado, D. R. Lamar, D. Schmucker, The molecular diversity of Dscam is functionally required for neuronal wiring specificity in *Drosophila*. *Cell* **125**, 607–620 (2006).
9. J. L. Lefebvre, D. Kostadinov, W. V. Chen, T. Maniatis, J. R. Sanes, Protocadherins mediate dendritic self-avoidance in the mammalian nervous system. *Nature* **488**, 517–521 (2012).
10. M. R. Sawaya, W. M. Wojtowicz, I. Andre, B. Qian, W. Wu, D. Baker, D. Eisenberg, S. L. Zipursky, A double S shape provides the structural basis for the extraordinary binding specificity of Dscam isoforms. *Cell* **134**, 1007–1018 (2008).
11. R. Meijers, R. Puettmann-Holgado, G. Skiniotis, J.-H. Liu, T. Walz, J.-H. Wang, D. Schmucker, Structural basis of Dscam isoform specificity. *Nature* **449**, 487–491 (2007).
12. S. E. Martinez, A. Y. Wu, N. A. Glavas, X.-B. Tang, S. Turley, W. G. J. Hol, J. A. Beavo, The two GAF domains in phosphodiesterase 2A have distinct roles in dimerization and in cGMP binding. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13260–13265 (2002).
13. J. Pandit, M. D. Forman, K. F. Fennell, K. S. Dillman, F. S. Menniti, Mechanism for the allosteric regulation of phosphodiesterase 2A deduced from the X-ray structure of a near full-length construct. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 18225–18230 (2009).
14. Y. Tang, M. A. Holbert, N. Delgoshaie, H. Wurtele, B. Guillemette, K. Meeth, H. Yuan, P. Drogaris, E. H. Lee, C. Durette, P. Thibault, A. Verreault, P. A. Cole, R. Marmorstein, Structure of the Rtt109-AcCoA/Vps75 complex and implications for chaperone-mediated histone acetylation. *Structure* **19**, 221–231 (2011).
15. D. Su, Q. Hu, H. Zhou, J. R. Thompson, R.-M. Xu, Z. Zhang, G. Mer, Structure and histone binding properties of the Vps75-Rtt109 chaperone-lysine acetyltransferase complex. *J. Biol. Chem.* **286**, 15625–15629 (2011).
16. L. H. Chao, M. M. Stratton, I. H. Lee, O. S. Rosenberg, J. Levitz, D. J. Mandell, T. Kortemme, J. T. Groves, H. Schulman, J. Kuriyan, A mechanism for tunable autoinhibition in the structure of a human Ca^{2+}/calmodulin-dependent kinase II holoenzyme. *Cell* **146**, 732–745 (2011).
17. P. Rellos, A. C. W. Pike, F. H. Niesen, E. Salah, W. H. Lee, F. von Delft, S. Knapp, Structure of the CaMKIIδ/calmodulin complex reveals the molecular mechanism of CaMKII kinase activation. *PLOS Biol.* **8**, e1000426 (2010).
18. Y. Gong, P. Cao, H.-J. Yu, T. Jiang, Crystal structure of the neurotrophin-3 and p75^{NTR} symmetrical complex. *Nature* **454**, 789–793 (2008).
19. X.-L. He, K. C. Garcia, Structure of nerve growth factor complexed with the shared neurotrophin receptor p75. *Science* **304**, 870–875 (2004).
20. W. Wu, G. Ahlsen, D. Baker, L. Shapiro, S. L. Zipursky, Complementary chimeric isoforms reveal Dscam1 binding specificity in vivo. *Neuron* **74**, 261–268 (2012).
21. G. Neves, J. Zucker, M. Daly, A. Chess, Stochastic yet biased expression of multiple *Dscam* splice variants by individual cells. *Nat. Genet.* **36**, 240–246 (2004).
22. G. Song, Y. Yang, J.-H. Liu, J. M. Casasnovas, M. Shimaoka, T. A. Springer, J.-H. Wang, An atomic resolution view of ICAM recognition in a complex between the binding domains of ICAM-3 and integrin $\alpha_L\beta_2$. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 3366–3371 (2005).
23. T. Schürpf, Q. Chen, J.-H. Liu, R. Wang, T. A. Springer, J.-H. Wang, The RGD finger of Del-1 is a unique structural feature critical for integrin binding. *FASEB J.* **26**, 3412–3420 (2012).
24. A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart, M. M. Etzler, *Essentials of Glycobiology* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, ed. 2, 2009).
25. T. J. Baribault, K. E. Neet, Effects of tunicamycin on NGF binding and neurite outgrowth in PC12 cells. *J. Neurosci. Res.* **14**, 49–60 (1985).
26. F. C. Neidhardt, H. E. Umbarger, Chemical composition of *Escherichia coli*, in *Escherichia coli and Salmonella: Cellular and Molecular Biology*, F. C. Neidhardt, Ed. (American Society of Microbiology Press, Washington, DC, ed. 2, vol. 1, 1996), p. 14.
27. E. A. Yamada, V. C. Sgarbieri, Yeast (*Saccharomyces cerevisiae*) protein concentrate: Preparation, chemical composition, and nutritional and functional properties. *J. Agric. Food Chem.* **53**, 3931–3936 (2005).
28. G. Apic, J. Gough, S. A. Teichmann, Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *J. Mol. Biol.* **310**, 311–325 (2001).
29. S.-A. Li, L. Cheng, Y. Yu, Q. Chen, Protein production, crystallization and preliminary X-ray analysis of two isoforms of the Dscam1 Ig7 domain. *Acta Crystallogr.* **F71**, 330–332 (2015).
30. L. Cheng, S.-A. Li, Y. Yu, Q. Chen, Protein production, crystallization and preliminary crystallographic analysis of the four N-terminal immunoglobulin domains of Down syndrome cell adhesion molecule 1. *Acta Crystallogr.* **F71**, 775–778 (2015).
31. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, PHENIX: A com-

Li *et al. Sci. Adv.* 2016;2:e1501118   27 May 2016

8 of 9

prehensive Python-based system for macromolecular structure solution. *Acta Crystallogr.* **D66**, 213–221 (2010).

32. P. Emsley, K. Cowtan, *Coot*: Model-building tools for molecular graphics. *Acta Crystallogr.* **D60**, 2126–2132 (2004).

33. V. B. Chen, W. B. Arendall III, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, *MolProbity*: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr.* **D66**, 12–21 (2010).