Cell, Volume *134*

A Double S Shape Provides the Structural Basis for the Extraordinary Binding Specificity of Dscam Isoforms

Michael R. Sawaya, Woj M. Wojtowicz, Ingemar Andre, Bin Qian, Wei Wu, David Baker, David Eisenberg, and S. Lawrence Zipursky

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Multiple Sequence Alignment

Amino acid sequences of the N-terminal eight Ig domains from 94 vertebrate and invertebrate *Dscam* genes were obtained from BLAST, Ensembl, FlyBase and from Brent Graveley (University of Connecticut). Sequences were aligned using MultAlin (Corpet, 1988) and gaps were removed. Alignments were visualized using WebLogo (Crooks et al., 2004; Schneider and Stephens, 1990).

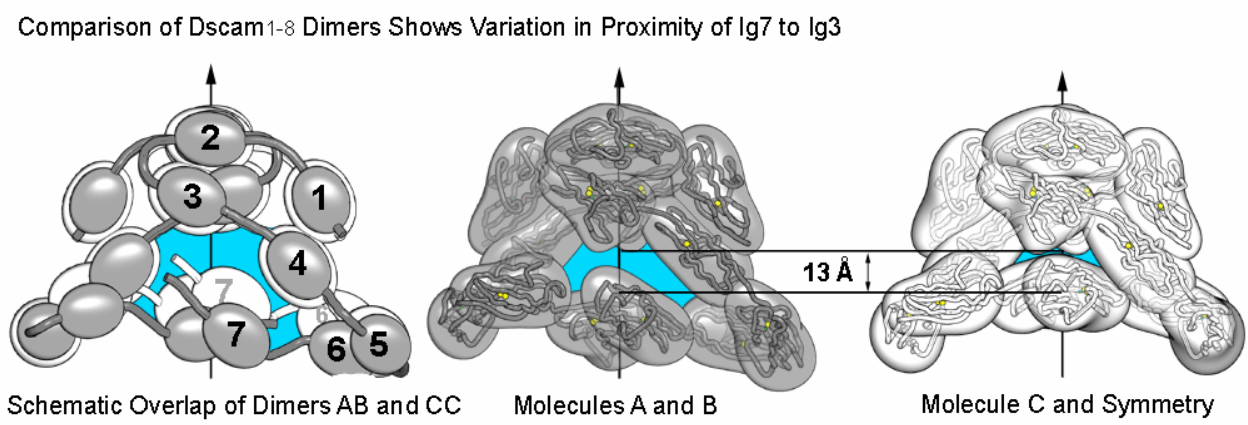

Figure S1. Hinge flexibility in Dscam$_{1-8}$

The dimer is oriented as in Figure 2. Molecules A and B are shown in gray ribbon. Molecule C and its symmetry mate are shown in white. The surfaces shown depict the molecular envelopes. Molecules are superimposed using αC atoms in Ig5 and Ig6. The largest difference in hinge angles are between Ig4 and Ig5 (26°) and between Ig6 and Ig7 (13°). The conformational changes produce a 13 Å shift that brings Ig7 closer to the horseshoe, Ig1-Ig4 in molecule C.

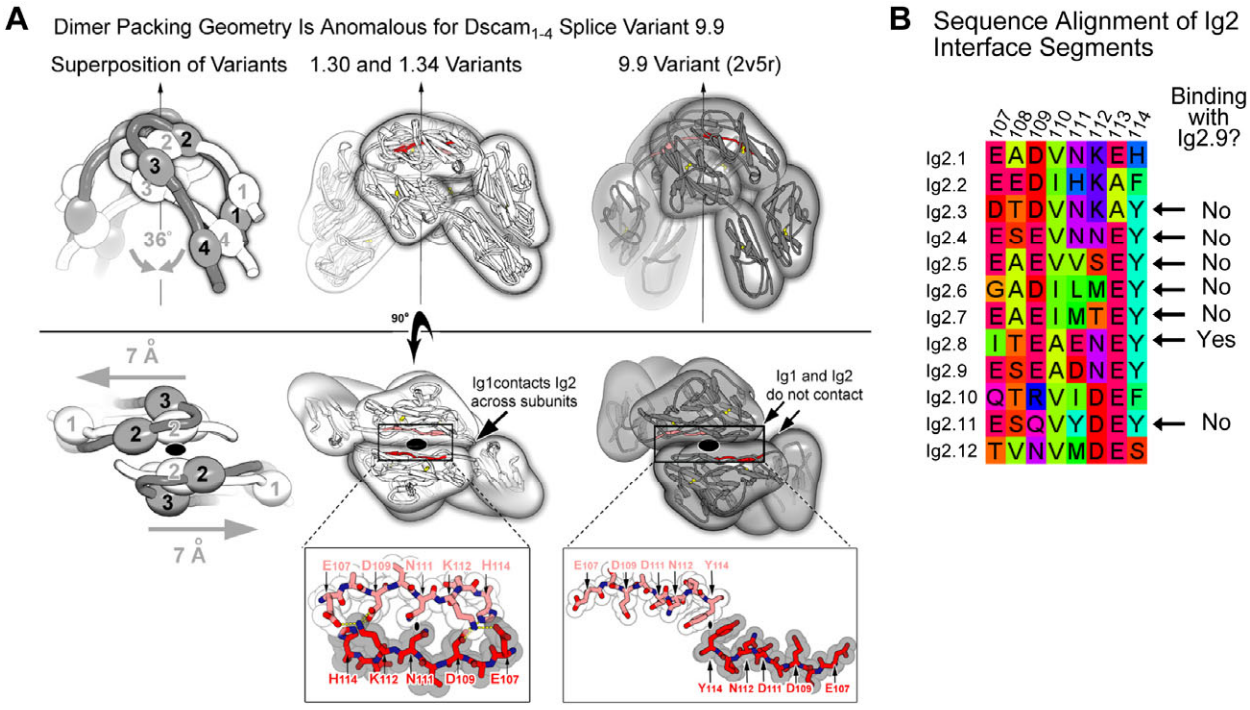

Figure S2. Ig1-Ig4 dimer geometry

The structure of the first four Ig domains of $Dscam_{1-8}$ is compared with the two $Dscam_{1-4}$ structures (Meijers et al., 2007). The two $Dscam_{1-4}$ structures contain different Ig2 and Ig3 variants. $Dscam_{1-4\ (1.34)}$ contains Ig2.1 and Ig3.34 and $Dscam_{1-4\ (9.9)}$ contains Ig2.9 and Ig3.9. The $Dscam_{1-4\ (9.9)}$ structure exhibits an alternate mode of dimerization. The alternate mode of dimerization in the $Dscam_{1-4\ (9.9)}$ structure is not consistent with the modular nature of variable domain interactions. Extensive binding studies on each of the three variable domains (Wojtowicz et al., 2007), the $Dscam_{1-4\ (1.34)}$ structure and the $Dscam_{1-8}$ structure reported here strongly argue that variable domain self-binding occurs in a modular fashion. Modularity provides the mechanism that underlies Dscam binding specificity because any combination of Ig2, Ig3 and Ig7 domains gives rise to a homophilic binding protein. Modularity requires that each of the three self-binding variable domains has a stereotyped interface that is unaffected by the identity of the other two variable domains. The lack of modularity in the $Dscam_{1-4\ (9.9)}$ dimerization interface is difficult to reconcile with the functional modularity observed. Evidence supporting the notion that the dimer interface observed in the $Dscam_{1-4\ (9.9)}$ structure represents a crystal packing artifact are outlined in the following panels.

(A). *Upper*. Superposition of Ig1-Ig4. *Left*. Schematic illustration of the Ig1-Ig4 dimer packing in $Dscam_{1-8}$ and $Dscam_{1-4\ (1.34)}$ (white) as compared with $Dscam_{1-4\ (9.9)}$ (gray). The molecules within the $Dscam_{1-4\ (9.9)}$ dimer exhibit a 36° rotation relative to the $Dscam_{1-4\ (1.34)}$ and $Dscam_{1-8}$ dimers. *Middle*. Superimposition of four Ig1-Ig4 dimers containing Ig2.1 reveals a conservation of the dimer interface. Dimer of molecules A/B and C/C from the $Dscam_{1-8}$ structure and two different structures of $Dscam_{1-4\ (1.34)}$ are superimposed. $Dscam_{1-8}$ and $Dscam_{1-4\ (1.34)}$ both contain Ig2.1 but differ at Ig3 ($Dscam_{1-8}$ contains Ig3.30 and $Dscam_{1-4\ (1.34)}$ contains Ig3.34). Despite the different Ig3 variants present in $Dscam_{1-8}$ and $Dscam_{1-4\ (1.34)}$, the Ig2-Ig2 interface is identical and the orientation of the Ig1-Ig4 molecules within the dimer is unchanged. *Right*. Structure of the $Dscam_{1-4\ (9.9)}$ dimer. The size of the $Dscam_{1-4\ (9.9)}$ dimerization interface (1,220 $Å^2$) is a small fraction of the area of the $Dscam_{1-4\ (1.34)}$ (3,711 $Å^2$) dimerization interface. *Lower*. Comparison of the Ig2.1-Ig2.1 and Ig2.9-Ig2.9 interfaces. The Ig2.1-Ig2.1 interface is composed of a symmetry axis (N111) flanked by left and right hydrogen-binding networks (see also Figure 4). Docking models of the remaining 11 Ig2 variants suggest that all of the Ig2 self-binding interfaces, including Ig2.9, adopt this same conformation with residue 111 at the symmetry axis (see Figure S4). The Ig2.9-Ig2.9 interface observed in the $Dscam_{1-4\ (9.9)}$ dimer is shifted such that residue Y114 resides at the symmetry axis. This shift may reflect the fact that the crystallization medium for $Dscam_{1-4\ (9.9)}$ lacked divalent ions. That divalent cations

may be an essential component of the biologically relevant Ig2.9-Ig2.9 interface is suggested by the proximity of E109 and D111 across the dimer interface in the Ig2.9 docking model. The cations are proposed to coordinate this E109-D111 pair. Crystallization at basic pH (pH 8.5) may have distorted the proper alignment of molecules by creating a strong electrostatic repulsion between E109-D111 pairs in the absence of divalent cations. In both the $Dscam_{1-8}$ and the $Dscam_{1-4\ (1.34)}$ structures *intra*molecular interactions between Ig1 and Ig2 position the Ig1 domains as bookends on either side of the Ig2 domains which guide the proper registration of the Ig2-Ig2 interface (see also Figure S3). These Ig1-Ig2 interactions are not present in the $Dscam_{1-4\ (9.9)}$ structure. Whether it is the lack of these Ig1-Ig2 interactions that leads to the shift at the Ig2.9-Ig2.9 interface or the shift at the Ig2.9-Ig2.9 interface that leads to the lack of interactions between Ig1 and Ig2 is not possible to determine. In addition, the shift observed in the $Dscam_{1-4\ (9.9)}$ dimer could be attributed to the fact that the Ig1-Ig4 construct was crystallized out of context of the larger Ig1-Ig8 segment known to be required for homophilic dimerization (Wojtowicz et al., 2004) and, as such, may have been susceptible to distortion by crystal packing forces. (B) Ig2 self-binding binding specificity. The 12 Ig2 variants exhibit striking binding specificity (Wojtowicz et al., 2007). Each binds to itself but rarely to other variants. A binding grid testing all pairwise combinations of Ig2 variants demonstrated that only 4 out of the 66 Ig2 pairs exhibit non-self binding (Ig2.1 and Ig2.3, Ig2.5 and Ig2.7, Ig2.6 and Ig2.7, Ig2.8 and Ig2.9). Swapping experiments demonstrated that the segment comprising residues 107-114 is sufficient to confer the binding specificity of all Ig2 variants. This region comprises the Ig2 dimerization interface observed in the $Dscam_{1-4\ (1.34)}$ and $Dscam_{1-8}$ structures as well as the docking models of the remaining 11 Ig2 variants (see Figure S4). The unique binding specificity of each Ig2 variant is proposed to arise from a unique combination of residues 107-114 which form left and right hydrogen-binding networks flanking the symmetry axis. The Ig2 dimerization interface in $Dscam_{1-4\ (9.9)}$ is composed of only a single residue (i.e. Y114) which resides at the symmetry axis. There are no flanking hydrogen-bond networks. If Y114 were sufficient to mediate the Ig2.9-Ig2.9 interface, then other Ig2 variants containing Y114 might be expected to bind to Ig2.9. Residues 107-114 for all Ig2 variants are shown. In addition to Ig2.9, seven Ig2 variants contain Y114. Of these, only one (i.e. Ig2.8) binds to Ig2.9. Therefore, the binding specificity of the Ig2 variants is inconsistent with the alternate mode of $Dscam_{1-4\ (9.9)}$ dimerization.

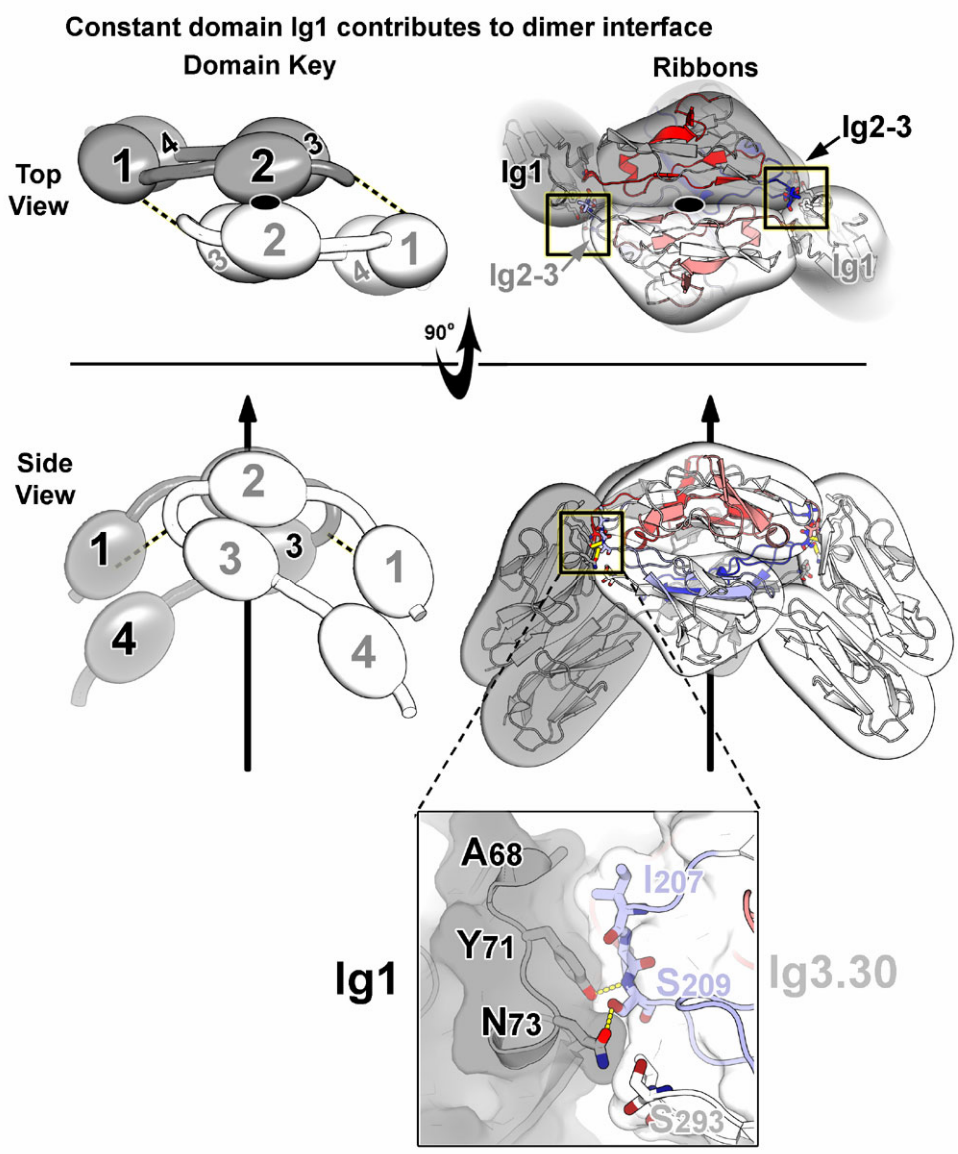

Figure S3. Ig1 domains "bookend" Ig2 domains in Dscam$_{1\text{-}8}$ dimer
Constant domain Ig1 contributes approximately 15% of the Dscam dimer interface surface area. Specifically, Ig1 shares hydrogen bonds and van der Waals contacts with the variable loop connecting Ig2 and Ig3 (lower inset). These contacts appear to act as a physical barrier to prevent the two molecules from sliding across a relatively flat interface (in the horizontal direction). The left panels provide a key for identifying the domains in the more detailed ribbon diagrams shown on the right. Only the first four Ig domains from Dscam$_{1\text{-}8}$ are illustrated. Gray and white ellipsoids distinguish the two molecules of the dimer. Red and blue colors distinguish the variable residues of Ig2 and Ig3, respectively. Black ellipses mark the axis of two-fold rotational symmetry.

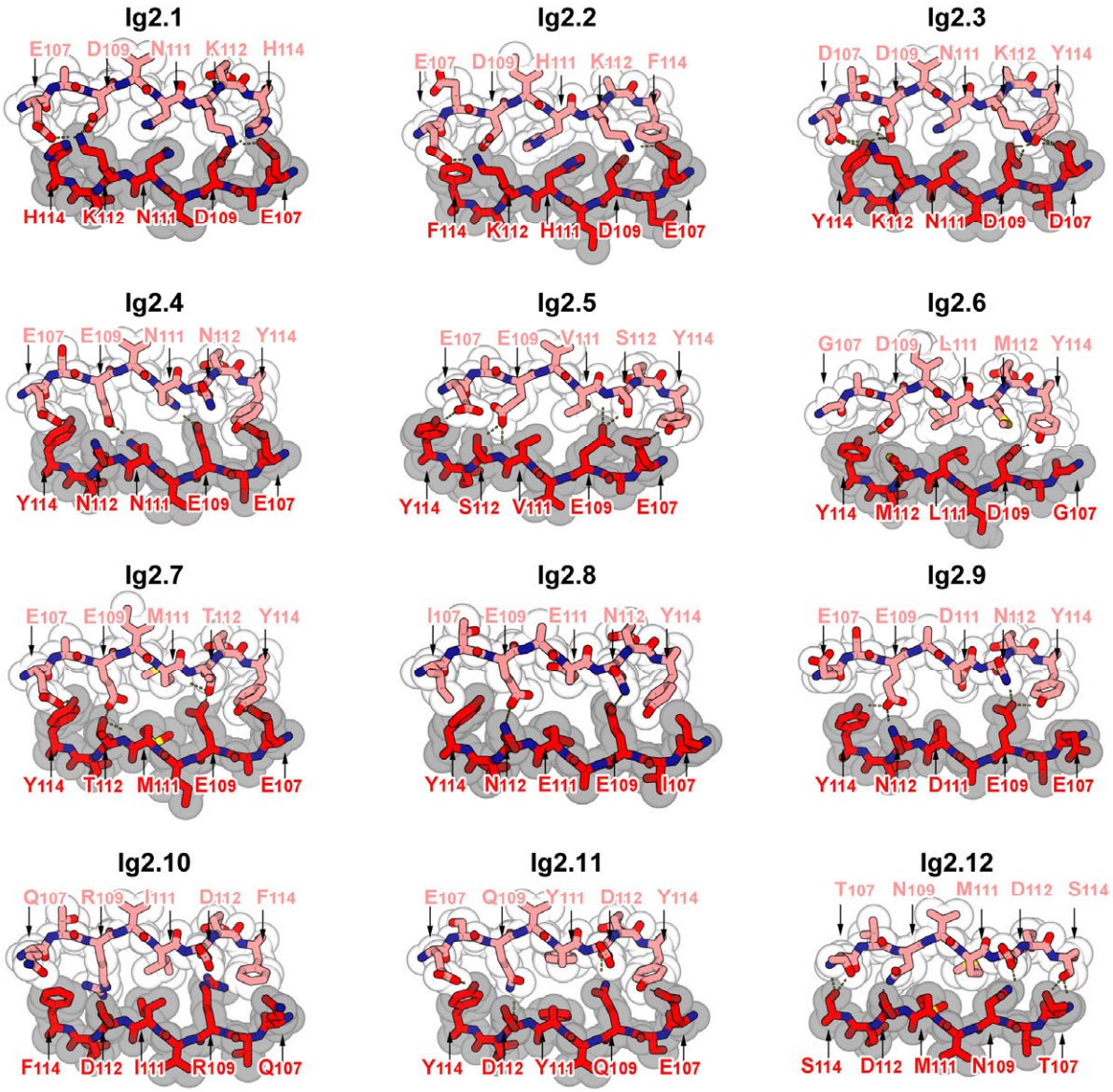

Figure S4. Ig2 homophilic interface docking models

Symmetric homodimeric docking models (André et al., 2007) were generated for each of the 12 Ig2 variable domains using Rosetta (Das et al., 2007). Similarity between the Ig2.1-Ig2.1 docking model interface and the Ig2.1-Ig2.1 interface in the $Dscam_{1\text{-}8}$ and previous $Dscam_{1\text{-}4\ (1.34)}$ crystal structures (Meijers et al., 2007) (see Figure 4A) validates the use of docking models to investigate the Ig2-Ig2 interface (this is also the case for Ig7 docking models, see Figure S6). The interface segment in the docking models of all 12 Ig2 variants comprises residues 107-114, the same region observed in the crystal structures of Ig2.1. That these segments reside at the interface of all Ig2 variants is strongly supported by previous biochemical swapping experiments which demonstrated that the 107-114 segment is sufficient to confer the binding specificity of all Ig2 variants (Wojtowicz et al., 2007). In all 12 docking models, residue 111 resides at the symmetry axis and is capable of packing against its symmetry mate. Ig2.8 and Ig2.9 which contain E111 and D111, respectively, likely require the presence of a cation to coordinate these negatively charged residues across the interface. Flanking the symmetry axis, each of the 12 Ig2 variants has a unique left and right network that exhibits electrostatic and shape complementarity. The electrostatic and shape complementarity exhibited at each interface provides a structural explanation for the highly-specific self-binding properties exhibited by each of the 12 Ig2 variants (Wojtowicz et al., 2007).

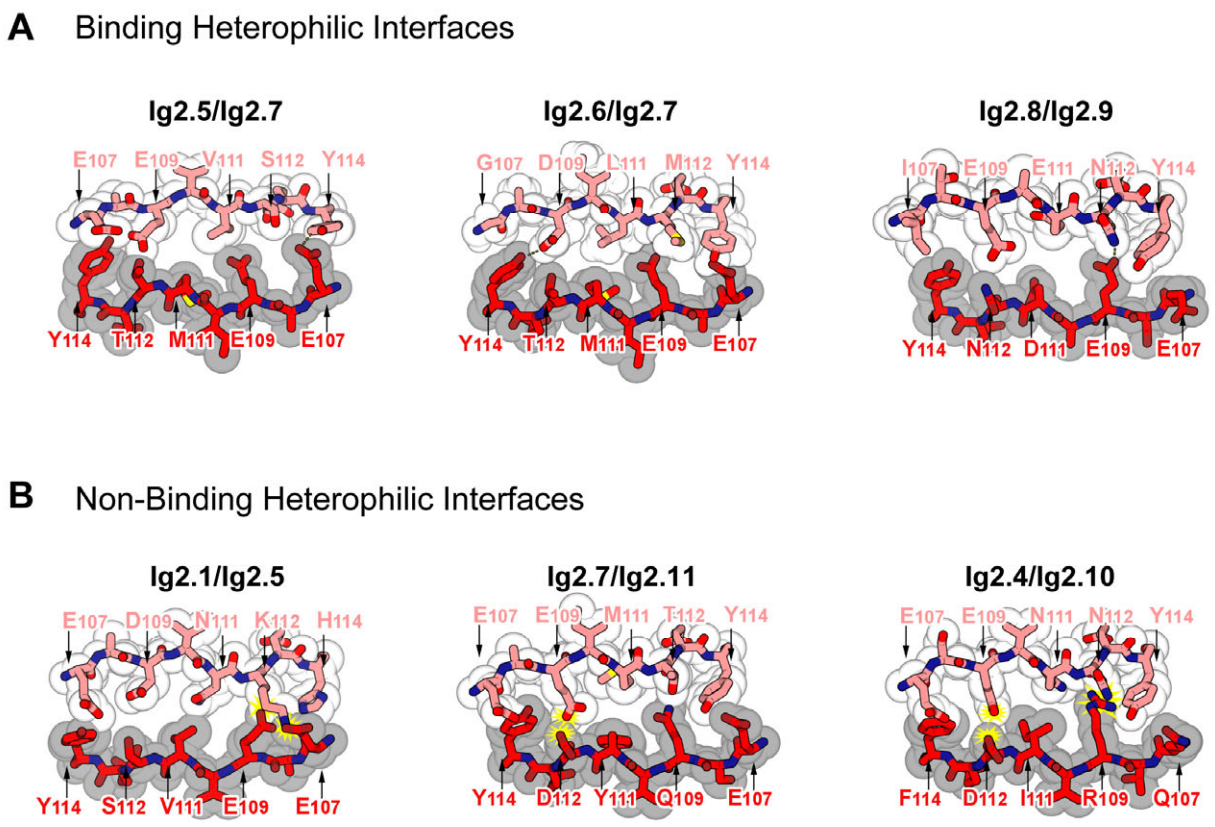

Figure S5. Ig2 heterophilic interface docking models

Symmetric homodimeric docking models (André et al., 2007) were generated for each of the 12 Ig2 variable domains using Rosetta (Das et al., 2007) (see Figure S4). Binding and non-binding Ig2 heterophilic interfaces are illustrated here.

(A). Binding Ig2 heterophilic interfaces. The 12 Ig2 domains were tested for binding in a grid fashion (i.e. 12 x 12) (Wojtowicz et al., 2007). Out of the 66 heterophilic pairs, 4 pairs exhibit heterophilic binding, albeit at levels lower than the homophilic binding of each. One binding Ig2 heterophilic pair is shown in Figure 4D and the other three pairs are shown here. All pairs exhibit electrostatic and shape complementarity, albeit to a lesser degree than each self-binding homophilic interface (compare with homophilic interfaces in Figure S4).

(B). Non-binding Ig2 heterophilic interfaces. Electrostatic and shape non-complementarity (yellow starbursts) are observed at either the left, right or both networks. Non-complementarity arises from steric overlap, electrostatic repulsion, and poor shape complementarity between mismatched side chains across the heterophilic interface.

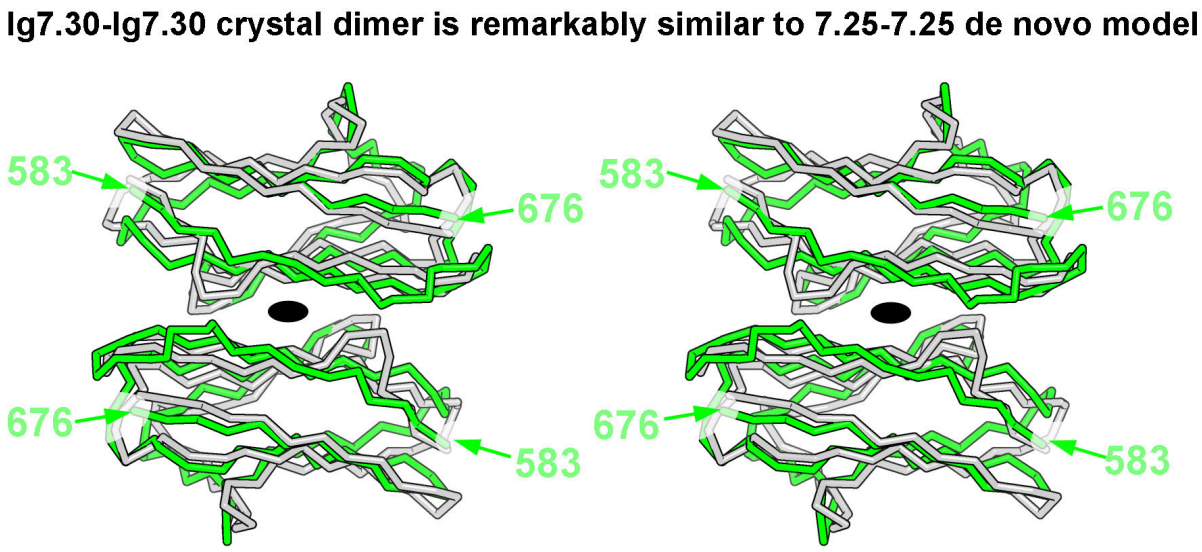

Figure S6. Superimposition of Ig7 dimer from crystal structure and de novo modeling
A remarkable similarity is observed between the Ig7.30-Ig7.30 dimer in the $Dscam_{1-8}$ crystal structure and the Ig7.25-Ig7.25 dimer model predicted using Rosetta (Wojtowicz et al., 2007). A stereo diagram of the superimposed Ig7.30-Ig7.30 and Ig7.25-g7.25 dimers is shown. The two dimers were superimposed intact, yielding an RMS deviation of 2.1 Å for 180 out of 190 possible alpha carbon pairs. The close coincidence in dimer geometry is remarkable given the difference in methods used to arrive at the models and the difference in amino acid sequence between the two isoforms – they share only 38% identity differing at 59 out of 95 residues. The green chain corresponds to the crystal structure. The gray chain corresponds to the de novo model.

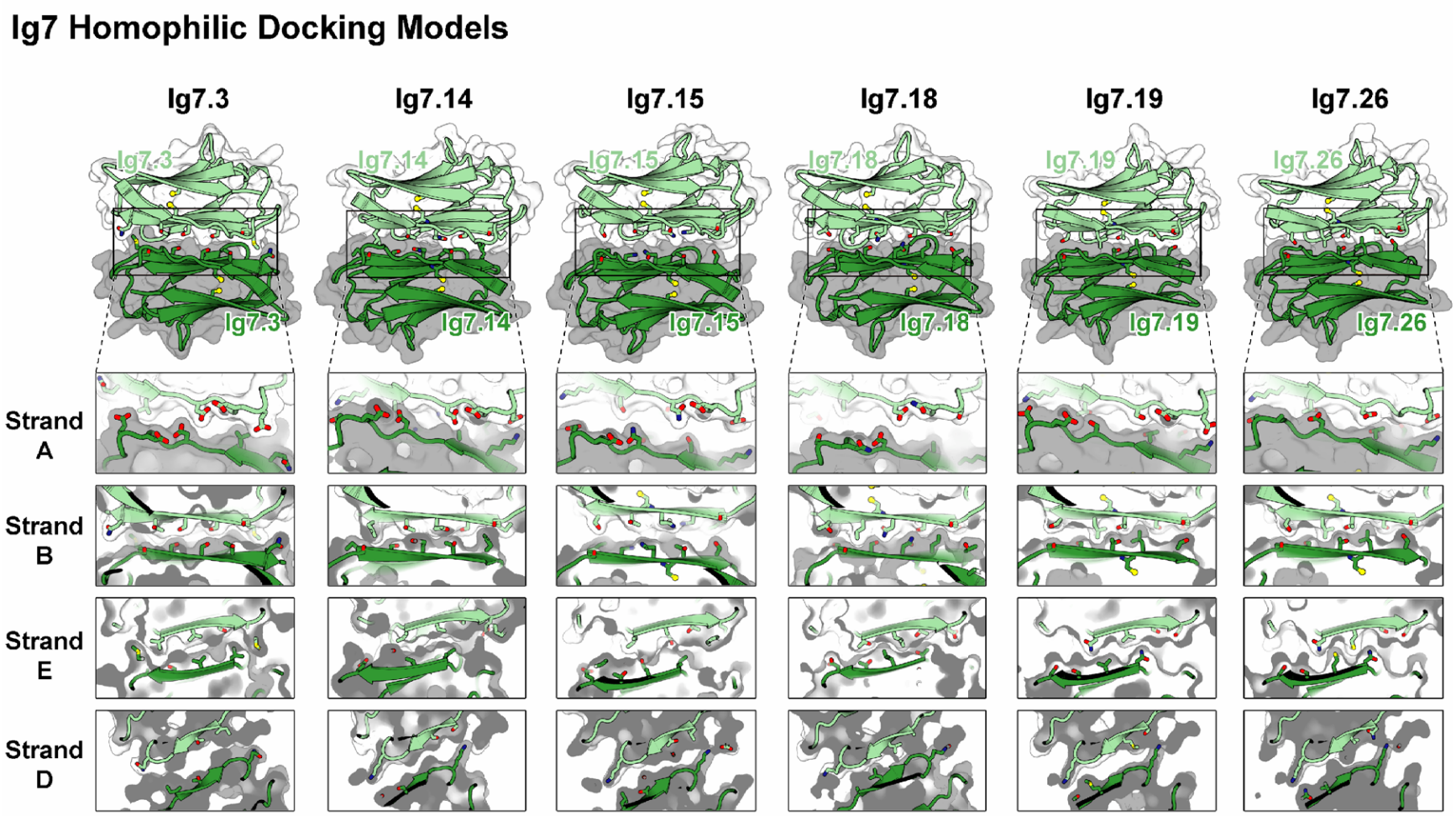

Figure S7. Ig7 Homophilic Docking Models

Homophilic docking models were generated for 10 Ig7 domains using Rosetta (Das et al., 2007; André, et al., 2007). Six docking models are shown here, one is shown in Figure 5B (i.e. Ig7.20) and three others are not shown. Molecule A is shown in lighter colors. All two-fold symmetric Ig7 interfaces are formed between identical segments along the ABED strands that comprise one face of the Ig7 domain. Each self-binding Ig7 interface exhibits complementarity across all of the ABED strands. Slices of each interface are taken from four different depths to illustrate surface complementarity between subunits.

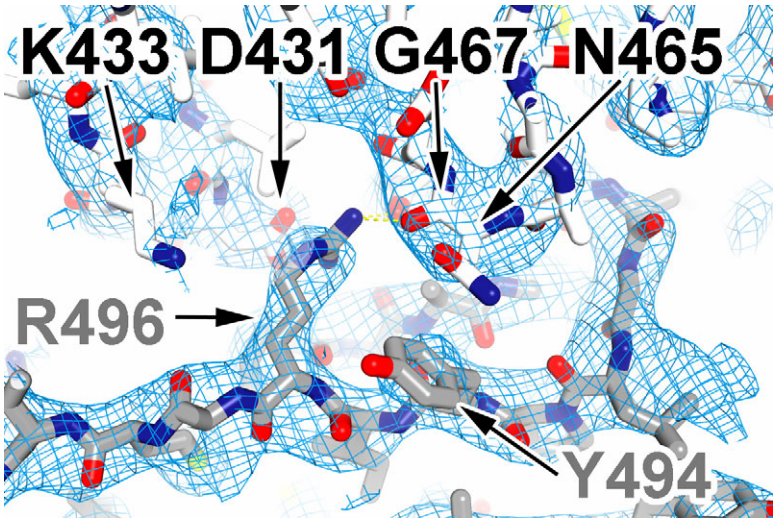

Figure S8. Electron density map of Ig5:Ig6 *intra*molecular interactions
The quality of the $Dscam_{1-8}$ model is illustrated by the features of this 4.2 Å resolution 2Fo-Fc omit map showing the *intra*molecular interactions between Ig5 and Ig6. All residues composing the Ig5:Ig6 hinge of molecules A, B, and C were removed from the coordinate set. The remaining coordinates were submitted to simulated annealing refinement (starting temperature 1500 K). The resulting coordinates were used to calculate the 2Fo-Fc omit map shown here. The map shows clearly defined density for R496 which appears to play a major role in supporting this hinge conformation.

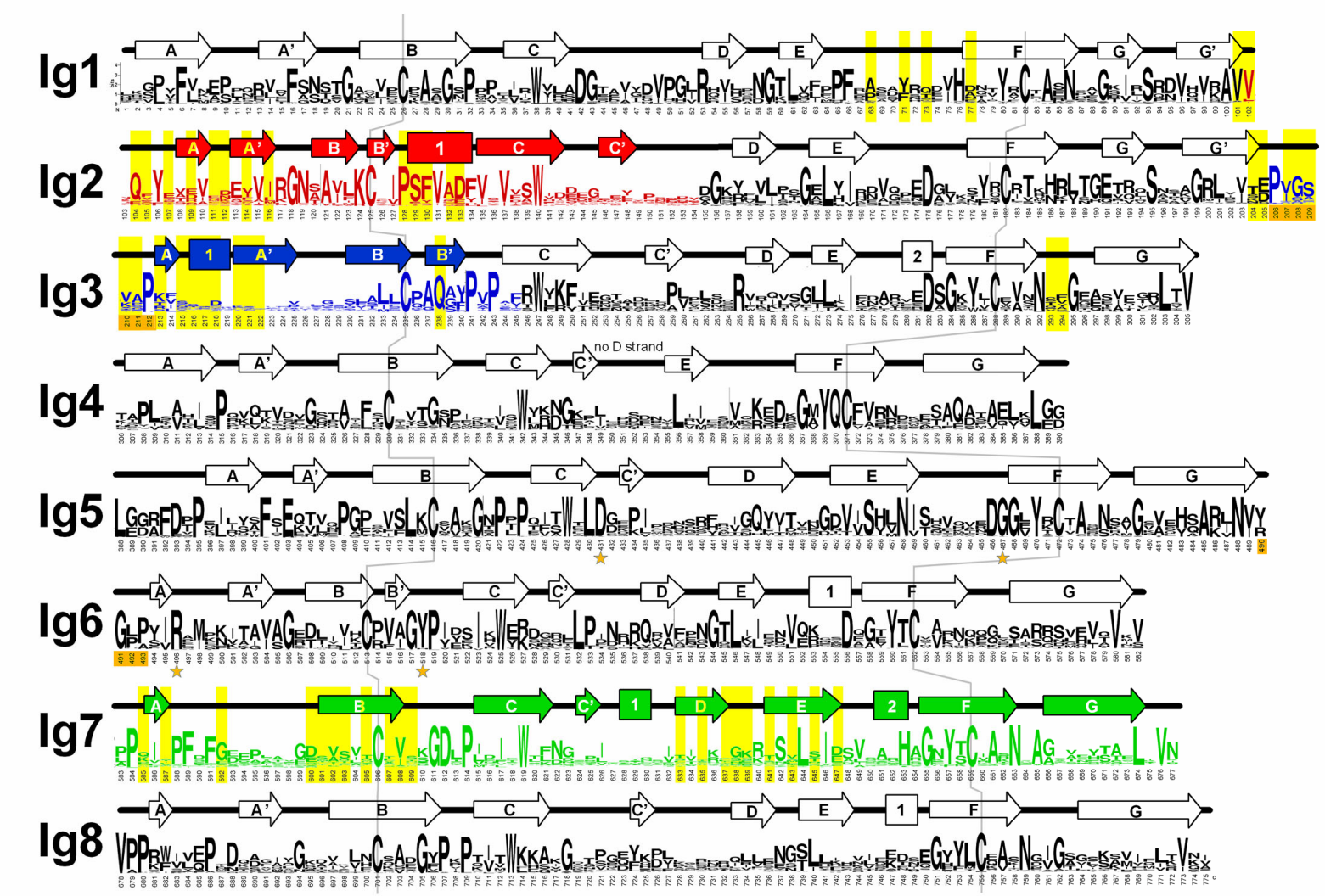

Figure S9. Multiple sequence alignment of 94 Dscam$_{1\text{-}8}$ proteins

Illustration of sequence conservation of the first 8 ectodomains of the Dscam molecule. Dscam amino acid sequences were collected from Dscam proteins including invertebrate Dscam, Dscam2, Dscam3 and Dscam4, and vertebrate DSCAM and DSCAM-L. In total, 94 different *Dscam* genes are represented including 295 exon 4s (encoding the N-terminal half of Ig2), 877 exon 6s (encoding the N-terminal half of Ig3) and 671 exon 9s (encoding Ig7). The size of each amino acid's character is proportional to the level of conservation. Red, blue, and green segments correspond to the invertebrate variable exons encoding Ig2, Ig3, and Ig7, respectively. The secondary structure is indicated above the sequence (β-strands, arrows; helices, rectangles). Yellow boxes indicate residues directly involved in dimer contacts. Note, these lie primarily in variable regions. Orange boxes indicate extended loops involved in producing reverse turns between domains, specifically between Ig2-Ig3, and Ig5-Ig6. Stars indicate especially conserved residues involved in the Ig5-Ig6 hinge. Sequences were aligned using Multalin (Corpet, 1988) and illustrated using WebLogo (http://weblogo.berkeley.edu/) (Crooks et al., 2004; Schneider and Stephens, 1990)

Table S1. Statistics of X-ray data collection and atomic refinement (numbers in parentheses refer to the outer shell of data).

Space group	*I*222
Unit cell parameters [Å,°]	*a*=118.6, *b*=177.9, *c*=434.3
Resolution range [Å]	90.0-4.2 (4.35-4.20)
R_{sym}* [%]	6.6 (46.2)
Number of unique data	34106(5605)
Completeness of data [%]	99.6 (99.9)
I/σ(*I*)	17.4 (4.0)
Number of residues (3 chains / asymmetric unit)	2127
Number of protein atoms	16435
Number of carbohydrate atoms	320
Number of glycerol and sulfate atoms	29
Number of water atoms	0
Matthews' coefficient [Å^3/Da]**	4.2
R [%]***	28.0 (29.1)
R_{free} [%]****	32.7 (36.8)
Test set size [%], selection	5.0, random
RMS deviations from target values	
Bond lengths [Å]	0.008
Bond angles [°]	1.2
Ramachandran angles	
Most favored [%]	84.2
Additionally allowed [%]	14.4
Generously allowed [%]	0.8
Disallowed [%]	0.7
Errat Overall Quality Factor[%]	78.8
Verify3D residues with score >0.2[%]	88.3
Average B factor for mainchain atoms [Å^2]	53.7
Average B factor for sidechain atoms [Å^2]	54.3
Average RMS B for mainchain atoms [Å^2]	0.56
Average RMS B for sidechain atoms [Å^2]	0.46

* $R_{sym} = \Sigma|I_o - I_o \text{ (mean)}|^2 / \Sigma [I_o^2]$, where I_o is the observed intensity. Both summations involve all input reflections for which more than one symmetry equivalent is averaged.
** Matthews' coefficient as defined by Matthews, 1968
*** $R = \Sigma||F_o| - |F_c||/ \Sigma|F_o|$
**** R_{free} as defined by Brünger, 1992

Table S2. *Intra*molecular Domain:Domain Interface Areas in Chain A

***Intra*molecular Domain Interfaces**	**Total Area Buried (Å^2)**
Ig1:Ig4	1351
Ig2:Ig3	1416
Ig4:Ig5	421
Ig5:Ig6	842
Ig6:Ig7	376
Ig7:Ig8	443

Table S3. Intermolecular Domain-Domain Interface Surface Areas

Interface	Reference	Chains	First Domain (Å^2)	Second Domain (Å^2)	Interface Area Total (Å^2)	Sc
Ig2-Ig2						
Ig2.1	this paper	A-B	546	550	1,096	0.612
Ig2.1	this paper	C-C	580	580	1,160	0.594
Ig2.1	2v5m	A-B	556	556	1,112	0.741
Ig2.1	2v5s	A-B	568	560	1,128	0.724
Ig2.9	2v5r	A-B	187	197	384	0.545
Ig3-Ig3						
Ig3.30	this paper	A-B	283	280	563	0.478
Ig3.30	this paper	C-C	293	293	586	0.607
Ig3.34	2v5m	A-B	352	352	704	0.834
Ig3.34	2v5s	A-B	354	356	710	0.830
Ig3.9	2v5r	A-B	84	85	169	0.657
Ig7-Ig7						
Ig7.30	this paper	A-B	623	631	1254	0.314
Ig7.30	this paper	C-C	691	691	1382	0.352

Sc=Surface complementarity as defined by Lawrence and Coleman (1993) J. Mol. Biol., 234, p946 - p950 and computed with the CCP4 program, SC.

Table S4. *Intra*molecular Domain:Domain Hinge Angle Differences Among Chains A, B, and C.

	Chains A-B	Chains A-C	Chains B-C	Average
Ig4:Ig5	33°	26°	17°	25°
Ig5:Ig6	0.5°	0.4°	0.4°	0.4°
Ig6:Ig7	12°	13°	13°	13°
Ig7:Ig8	Flexibility in this hinge is so great that Ig8 is disordered in chains A and C.			

Calculated using LSQKAB from CCP4 suite of crystallographic programs. Collaborative Computational Project, Number 4. (1994). The CCP4 suite: programs for protein crystallography. Acta Cryst. D50, 760-763.

SUPPLEMENTAL REFERENCES

André, I., Bradley, P., Wang, C., Baker, D. (2007). Prediction of the structure of symmetrical protein assemblies. Proc. Natl. Acad. Sci. USA *104*, 17656-17661.

Brünger, A. T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature *355*, 472-474.

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res *16*, 10881-10890.

Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. Genome Res *14*, 1188-1190.

Das, R., Qian, B., Raman, S., Vernon, R., Thompson, J., Bradley, P., Khare, S., Tyka, M. D., Bhat, D., Chivian, D., Kim, D. E., Sheffler, W. H., Malmström, L., Wollacott, A. M., Wang, C., Andre, I., Baker, D. (2007). Structure prediction for CASP7 targets using extensive all-atom refinement with Rosetta@home. Proteins *69 Suppl 8*, 118-128.

Matthews, B. W. (1968). Solvent content of protein crystals. J. Mol. Biol. *33*, 491-497.

Meijers, R., Puettmann-Holgado, R., Skiniotis, G., Liu, J. H., Walz, T., Wang, J. H., and Schmucker, D. (2007). Structural Basis of Dscam Isoform Specificity. Nature *449*, 487-491.

Schneider, T. D., and Stephens, R. M. (1990). Sequence logos: a new way to display consensus sequences. Nucleic Acids Res *18*, 6097-6100.

Wojtowicz, W. M., Flanagan, J. J., Millard, S. S., Zipursky, S. L., and Clemens, J. C. (2004). Alternative splicing of Drosophila Dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding. Cell *118*, 619-633.

Wojtowicz, W. M., Wu, W., Andre, I., Qian, B., Baker, D., and Zipursky, S. L. (2007). A vast repertoire of Dscam binding specificities arises from modular interactions of variable Ig domains. Cell *130*, 1134-1145.