

Dissertation  
submitted to the  
Combined Faculties for the Natural Sciences and for Mathematics  
of the Ruperto-Carola University of Heidelberg, Germany  
for the degree of  
Doctor of Natural Sciences

presented by

Diplom- Chem. Magda Chegkazi

born in: Athens, Greece

Oral-examination: 09/10/2017



Structural characterization of cell surface receptors  
involved in axon guidance

Referees: Dr. Thomas R. Schneider  
Prof. Dr. Ilme Schlichting



*“There are no facts, only interpretations”*

*Friedrich Nietzsche*



# CONTENTS

List of Abbreviations.....	9
List of Figures.....	10
List of Tables .....	11
SUMMARY .....	15
ZUSAMMENFASSUNG .....	17
Chapter 1 .....	21
INTRODUCTION .....	21
1.1 Physiology of the neuron .....	21
1.2 Neuronal wiring .....	22
1.3 Axon guidance for neuronal wiring .....	25
1.4 Netrin as a guidance cue .....	27
1.4.1 Cell receptors of netrin-1 .....	28
1.5 Deleted in Colorectal Cancer .....	29
1.6 Down Syndrome Cell Adhesion Molecule .....	31
1.6.1 Dscam in <i>Drosophila melanogaster</i> .....	32
1.6.2 DSCAM in vertebrates .....	35
1.7 Aim of the study .....	39
Chapter 2 .....	41
MATERIALS AND METHODS.....	41
2.1 Materials .....	41
2.2 Methods .....	41
2.2.1 Constructs .....	41
2.2.2 Vector design.....	42
2.2.3 Molecular cloning.....	43
2.2.3.1 Preparation of insert gene .....	43
2.2.3.2 DNA digestion and ligation in restriction endonuclease cloning ....	44
2.2.3.3 Overlap extension cloning .....	45
2.2.3.4 Agarose gel electrophoresis .....	46
2.2.3.5 Transformation of <i>E. coli</i> competent cells.....	46
2.2.3.6 Plasmid preparation.....	47
2.2.3.7 Confirmation of construct.....	47
2.2.3.8 Culture preparation for plasmid amplification .....	47
2.2.4 Protein expression in Human Embryonic Kidney cells .....	48
2.2.4.1 Preparation of HEK cells .....	48
2.2.4.2 Cell maintenance .....	48
2.2.4.3 Protein transfection and large-scale expression.....	49
2.2.5 Protein purification .....	50
2.2.6 Protein detection and purity assessment .....	53
2.2.6.1 SDS-PAGE .....	53

2.2.6.2 Western blot .....	54
2.2.7 Biophysical methods .....	55
2.2.7.1 MALDI-TOF analysis .....	55
2.2.7.2 Thermal shift assay (Thermofluor).....	55
2.2.7.3 FPLC-RALS/RI/UV .....	56
2.2.7.4 Isothermal titration calorimetry .....	57
2.2.8 Structural characterization.....	59
2.2.8.1 Small-angle X-ray scattering.....	59
2.2.8.2 Electron microscopy .....	61
2.2.8.3 X-ray protein crystallography.....	63
<b>Chapter 3 .....</b>	<b>73</b>
<b>RESULTS AND DISCUSSION.....</b>	<b>73</b>
<b>3.1 Results .....</b>	<b>73</b>
3.1.1 Construct design rationale of DSCAM Ig1-Ig8 & Ig1-Ig9 .....	74
3.1.2 Purification&biophysical characterization of DSCAM Ig1-Ig8&Ig1-Ig9 ..	75
3.1.3 Structural characterization of DSCAM Ig1-Ig8 & Ig1-Ig9 .....	80
3.1.3.1 Small-angle X-ray scattering of DSCAM Ig1-Ig8 & Ig1-Ig9 .....	80
3.1.3.2 Electron microscopy of DSCAM Ig1-Ig8 & Ig1-Ig9 .....	84
3.1.3.3 X-ray crystallography of DSCAM Ig1-Ig8 & Ig1-Ig9 .....	93
3.1.4 Binding affinity studies and <i>in vitro</i> reconstitution of DSCAM Ig1-Ig9 WT:netrin <sub>VIV</sub> .....	99
3.1.5 Structural characterization of DSCAM Ig1-Ig9 WT:netrin <sub>VIV</sub> .....	100
3.1.5.1 Small-angle X-ray scattering of DSCAM Ig1-Ig9 WT:netrin <sub>VIV</sub> .....	100
3.1.5.2 Electron microscopy of DSCAM Ig1-Ig9 WT:netrin <sub>VIV</sub> .....	103
3.1.6 Construct design rationale of DSCAM Ig1-Ig4 .....	107
3.1.7 Purification & biophysical characterization of DSCAM Ig1-Ig4 .....	109
3.1.8 Structural characterization of DSCAM Ig1-Ig4 .....	111
3.1.8.1 Small-angle X-ray scattering of DSCAM Ig1-Ig4 .....	111
3.1.8.2 X-ray crystallography of DSCAM Ig1-Ig4.....	113
i. Structure determination and refinement of DSCAM Ig1-Ig4 WT .....	121
ii. DSCAM Ig1-Ig4 WT structural analysis.....	123
<b>3.2 Discussion .....</b>	<b>131</b>
<b>Chapter 4 .....</b>	<b>139</b>
<b>CONCLUSIONS AND PERSPECTIVES .....</b>	<b>139</b>
<b>REFERENCES .....</b>	<b>143</b>
<b>APPENDIX .....</b>	<b>157</b>
<b>ACKNOWLEDGEMENTS .....</b>	<b>165</b>



List of Abbreviations in alphabetical order.

Abbreviations	Name
a.a.	Amino acids
ABC	Dimethylamine-borane complex
APS	Ammonium persulfate
ASU	Asymmetric unit
Bis-Tris methane	2-[Bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol
BSA	Bovine serum albumin
BSA	Buried surface area
CAM	Cell adhesion molecule
CNS	Central nervous system
CV	Column volume
Da	Dalton
DCC	Deleted in colorectal cancer
DMEM	Dulbecco's modified eagle's medium
DLS	Dynamic light scattering
dNTPs	Deoxyribonucleotide triphosphates
DSCAM	Down Syndrome Cell Adhesion Molecule in vertebrates
Dscam	Down syndrome cell adhesion molecule in invertebrates
DTT	Dithiothreitol
$\epsilon$	Extinction coefficient
EC	Extracellular cadherin
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
EM	Electron microscopy
EndoH	Endoglycosidase H
$f_{bu}$	Fraction of fully buried atoms
$f_{np}B$	Non-polar interface area
FCS	Fetal calf serum
FN	Fibronectin
GnTI	<i>N</i> -acetylglucosaminyltransferase-I
GlcNAc	<i>N</i> -Acetyl-D-glucosamine
HEK	Human embryonic kidney
Hepes	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HTX	High-throughput crystallization
IEC	Ion exchange chromatography
Ig	Immunoglobulin
IMAC	Immobilized metal affinity chromatography
ITC	Isothermal titration calorimetry
LB	Lysogeny broth
MAD	Multiple-wavelength anomalous dispersion
MALDI-TOF	Matrix-assisted laser desorption/ionization-time of flight
Man	Mannose
MES	2-( <i>N</i> -Morpholino)-ethanesulfonic acid
MIR	Multiple isomorphous replacement
MIRAS	Multiple isomorphous replacement with anomalous scattering
MOPS	3-( <i>N</i> -morpholino)-propanesulfonic acid
MR	Molecular replacement
MS	Mass spectrometry
MW	Molecular weight
NCS	Non-crystallographic symmetry
NMR	Nuclear magnetic resonance

NSD	Normalized spatial discrepancy
OEC	Overlap extension cloning
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PCT	Pre-crystallization test
PDB	Protein data bank
PEG	Polyethylene glycol
PEI	Polyethylenimine
PGA-LM	Poly- $\gamma$ -glutamic acid of low molecular weight
PNS	Peripheral nervous system
RALS	Right-angle light scattering
RI	Refractive index
r.m.s.d.	Root-mean-square deviation
Robo	Roundabout
RT	Room temperature
SAD	Single-wavelength anomalous dispersion
SAXS	Small-angle X-ray scattering
S <sub>c</sub>	Shape complementarity
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SIR	Single isomorphous replacement
SIRAS	Single isomorphous replacement with anomalous scattering
SOC	Super optimal broth for catabolite repression
SPR	Surface plasmon resonance
TAE	Tris-acetate-EDTA
TCEP	Tris-(2-carboxyethyl)-phosphine
TEMED	Tetramethylethylenediamine
TEV	Tobacco etch virus
Tris base	2-Amino-2-(hydroxymethyl)-1,3-propanediol
Tris-HCl	Tris-(hydroxymethyl)-aminomethane hydrochloride
UNC	Uncoordinated
WB	Western blot

#### List of Figures.

Figure I-1. Schematic representation of a typical vertebrate neuron cell .....	22
Figure I-2. Schematic representation of the growth cone .....	23
Figure I-3. The diversity of axon guidance .....	24
Figure I-4. Schematic representation of the neural tube in the developing spinal cord in vertebrates .....	26
Figure I-5. Schematic representation of netrin family .....	28
Figure I-6. Schematic representation of cell surface receptors interacting with netrin-1 .....	29
Figure I-7. Schematic representation showing the biofunctional role of guidance cues in axon guidance .....	31
Figure I-8. Alternative m-RNA splicing of Dscam gene in <i>D. melanogaster</i> .....	33
Figure I-9. Schematic representation of the proposed model for Dscam homophilic interactions based on the crystal structure of Dscam Ig1-Ig8 in <i>Drosophila</i> .....	34
Figure I-10. Schematic representation of DSCAM in humans & Dscam in <i>D. melanogaster</i> ..	37
Figure M-1. Schematic representation of the pXLG vector .....	42
Figure M-2. Schematic representation of the pXLG vector with DSCAM .....	43
Figure M-3. Endoglycosidase H cleavage sites .....	67

Figure R-1. Schematic representation of protein constructs for DSCAM and netrin-1 .....	75
Figure R-2. Western blot from large-scale expression of DSCAM Ig1-Ig8 & DSCAM Ig1-Ig9. 76	
Figure R-3. SEC profile of DSCAM Ig1-Ig8 WT & GnTI <sup>-</sup> and SDS-PAGE after IMAC and SEC purifications .....	76
Figure R-4. DSCAM Ig1-Ig8 WT oligomeric state analysis with FPLC coupled with SLS detectors, using Malvern's OmniSEC GPC/SEC system .....	77
Figure R-5. DSCAM Ig1-Ig9 WT and GnTI <sup>-</sup> purification profiles after SEC and SDS-PAGE after IMAC and SEC purification .....	78
Figure R-6. Selected graphs of thermofluor assays for DSCAM Ig1-Ig8 WT .....	79
Figure R-7. Selected graphs of thermofluor assays for DSCAM Ig1-Ig9 WT .....	80
Figure R-8. SAXS analysis and model generation of DSCAM Ig1-Ig8 WT .....	81
Figure R-9. DSCAM Ig1-Ig8 GnTI <sup>-</sup> SAXS experimental data in solution in comparison to Dscam crystallographic structure from <i>Drosophila</i> .....	82
Figure R-10. Cluster of DSCAM Ig1-Ig9 rigid body models obtained by SAXS .....	83
Figure R-11. DSCAM Ig1-Ig8 WT and GnTI <sup>-</sup> raw images after negative staining .....	84
Figure R-12. Single-particle analysis of DSCAM Ig1-Ig9 WT .....	86
Figure R-13. Schematic representation of DSCAM Ig1-Ig9 in different conformations and orientations after two-dimensional reference-free classification .....	87
Figure R-14. Sequence alignment of DSCAM Ig1-Ig9 with protein receptors comprising Ig domains and hinge regions.....	89
Figure R-15. DSCAM Ig1-Ig9 WT and GnTI <sup>-</sup> crosslinking.....	91
Figure R-16. DSCAM Ig1-Ig9 WT crosslinking .....	92
Figure R-17. Crystallization and optimization of DSCAM Ig1-Ig8 GnTI <sup>-</sup> .....	94
Figure R-18. EndoH treatment of DSCAM Ig1-Ig8 GnTI <sup>-</sup> .....	96
Figure R-19. Crystallization of DSCAM Ig1-Ig9 WT and GnTI <sup>-</sup> .....	97
Figure R-20. Co-purification profile of DSCAM Ig1-Ig9 WT with netrin <sub>IV</sub> after SEC.....	99
Figure R-21. Netrin <sub>IV</sub> analysis derived from SAXS experimental data .....	100
Figure R-22. DSCAM Ig1-Ig9 WT:netrin <sub>IV</sub> rigid body models in complex formation .....	102
Figure R-23. Single-particle analysis of DSCAM Ig1-Ig9 WT with netrin <sub>IV</sub> in 1:1 molar ratio .....	105
Figure R-24. Single-particle analysis of DSCAM Ig1-Ig9 WT with netrin <sub>IV</sub> in 1:2 molar ratio .....	106
Figure R-25. Sequence alignment of human DSCAM Ig1-Ig4 with Dscam in <i>D. melanogaster</i> , DCC, axonin1, contactin4 and neurofascin .....	108
Figure R-26. Superposition of Dscam Ig1-Ig4 structure, in monomer, with structural homologous proteins adopting a horseshoe-like configuration .....	109
Figure R-27. DSCAM Ig1-Ig4 WT and GnTI <sup>-</sup> expression and purification profile .....	110
Figure R-28. SEC coupled with SLS detector of DSCAM Ig1-Ig4 WT .....	111
Figure R-29. SAXS analysis and model generation of DSCAM Ig1-Ig4 WT .....	112
Figure R-30. Crystallization and optimization of DSCAM Ig1-Ig4 GnTI <sup>-</sup> and WT .....	114
Figure R-31. DSCAM Ig1-Ig4 WT helical data collection at ID29 .....	116
Figure R-32. ILI test for twinning for data collected in space groups C222 <sub>1</sub> and P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub> ....	118
Figure R-33. The crystal structure of human DSCAM Ig1-Ig4 determined at 2.35 Å .....	125
Figure R-34. Ribbon representation of DSCAM Ig1-Ig4 molecules in the crystal .....	128
Figure R-35. Ribbon representation of Dscam in <i>Drosophila</i> in the ASU and of human DSCAM Ig1-Ig4 structure with a symmetry related molecule .....	129
Figure S1. Sequence alignment of DSCAM Ig1-Ig4 from <i>H. sapiens</i> with Dscam Ig1-Ig4 different isoforms from <i>D. melanogaster</i> .....	163

#### List of Tables.

Table I-1. Cell adhesion and axon guidance molecules with their receptors .....	25
Table M-1. List of constructs for protein expression in Human Embryonic Kidney cell line ...	41
Table M-2. Selected web tools for protein construct design and structure prediction .....	42
Table M-3. Primers for preparation of the insert gene .....	43
Table M-4. PCR reaction protocol for plasmid amplification .....	44

Table M-5. PCR thermocycling conditions for plasmid amplification .....	44
Table M-6. Restriction enzyme digestion reaction for inserts and vectors .....	45
Table M-7. Restriction enzyme ligation reaction for inserts with vectors .....	45
Table M-8. Composition of main buffers used for IMAC purification of DSCAM .....	50
Table M-9. Composition of main buffers used for SEC purification of DSCAM .....	51
Table M-10. Composition of buffers used for IEC purification of DSCAM Ig1-Ig9 .....	52
Table M-11. Composition of Bis-Tris gels .....	53
Table M-12. Composition of running buffers used for SDS-PAGE .....	53
Table M-13. Composition of buffers used for Western blot .....	54
Table M-14. Thermofluor 96-well assay .....	56
Table M-15. ITC experimental set-up and buffer composition .....	58
Table M-16. Protein crosslinking of DSCAM Ig1-Ig9 in PD-10 columns .....	63
Table M-17. Deglycosylation of DSCAM Ig1-Ig8 GnTI <sup>1</sup> with EndoH .....	68
Table M-18. X-ray absorption edges .....	70
Table R-1. Table of selected statistics for DSCAM Ig1-Ig8 and DSCAM Ig1-Ig9 derived from SAXS data collection .....	83
Table R-2. Data collection statistics of DSCAM Ig1-Ig8 GnTI <sup>1</sup> .....	95
Table R-3. Table of selected statistics for DSCAM Ig1-Ig9, netrin <sub>VIV</sub> & DSCAM Ig1-Ig9:netrin <sub>VIV</sub> derived from SAXS data collection .....	103
Table R-4. Ranking of DSCAM Ig1-Ig9:netrin <sub>VIV</sub> models for binding site 1&binding site 2 ...	103
Table R-5. Table of selected statistics for DSCAM Ig1-Ig4 derived from SAXS data collection .....	113
Table R-6. Summary of the crystallization methods and optimization approaches employed for all DSCAM constructs .....	115
Table R-7. Data collection statistics of DSCAM Ig1-Ig4 WT on C222 <sub>1</sub> .....	116
Table R-8. Data collection statistics of DSCAM Ig1-Ig4 WT on P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub> .....	117
Table R-9. Matthews coefficient as calculated by CCP4 suite for DSCAM Ig1-Ig4 WT .....	118
Table R-10. Results obtained from MR using Phaser for DSCAM Ig1-Ig4 WT .....	119
Table R-11. Refinement statistics for DSCAM Ig1-Ig4 WT by BALBES .....	119
Table R-12. Diffraction data statistics of DSCAM Ig1-Ig4 WT soaked with YbCl <sub>3</sub> .H <sub>2</sub> O .....	120
Table R-13. Diffraction data statistics of DSCAM Ig1-Ig4 WT soaked with K <sub>2</sub> PtCl <sub>4</sub> .....	120
Table R-14. Data collection and refinement statistics of DSCAM Ig1-Ig4 WT on C222 <sub>1</sub> .....	122
Table R-15. Buried surface area and number of residues (Nres) involved in interdomain interactions of Dscam Ig1-Ig4 homodimers in comparison to DSCAM interface involving the same domains .....	130
Table D-1. Summary of methods employed and results obtained for human DSCAM .....	132
Table S1. List of laboratory consumables and equipment in alphabetical order .....	157
Table S2. List of chemicals in alphabetical order .....	158
Table S3. List of antibodies in alphabetical order .....	159
Table S4. List of enzymes in alphabetical order .....	159
Table S5. List of primers for construct design .....	159
Table S6. SAXS data collection and scattering parameters for DSCAM Ig1-Ig8 WT .....	160
Table S7. SAXS data collection and scattering parameters for DSCAM Ig1-Ig9 WT .....	160
Table S8. SAXS data collection and scattering parameters for DSCAM Ig1-Ig9:netrin <sub>VIV</sub> .....	161
Table S9. SAXS data collection and scattering parameters for DSCAM Ig1-Ig4 WT .....	161
Table S10. List of amino acids .....	162





## SUMMARY

During neuron development, the establishment of precise neuron connections is crucial for the correct wiring pattern formation. The effectiveness of such a neuronal network is secured by repulsive signals, generated upon homophilic interactions between molecular probes that prevent synapse formation with sister dendrites (self-avoidance). In addition, the presence of surface receptors on the cells dictates the axon orientation and growth, in response to signaling molecules (guidance cues) through attractive or repulsive mechanisms.

Down Syndrome Cell Adhesion Molecule (DSCAM) is a cell surface receptor, the ectodomain of which, comprises 10 Immunoglobulin (Ig-like) and six Fibronectin (FNIII) domains. It is expressed, mainly, in commissural axons in mammals and is known to control neuronal response through heterophilic binding to guidance cues, like netrin-1. Their interaction induces axon chemoattraction that triggers cytoplasmic signaling, leading to axonal growth. In addition, DSCAM mediates homophilic interactions, essential for promoting cell adhesion and aggregation. Its homologue in *D. melanogaster* (Dscam), upon homophilic binding, has a different role and appears to be involved, mainly, in neuron self-avoidance and dendritic discrimination. The aim of the thesis was to identify the domains involved in homodimerization of human DSCAM and explore its heterophilic interactions with netrin-1, in order to assess the mechanistic differences between the two species. To this end, biophysical and structural studies involving SAXS, EM and X-ray crystallography were performed.

A selection of DSCAM constructs comprising the *N*-terminal domains Ig1-Ig4, Ig1-Ig8 and Ig1-Ig9 was designed, successfully expressed and isolated in high purity. The results obtained from biophysical characterization assays and SAXS, suggested that DSCAM homodimerization probably follows a different mechanism than the one observed in Dscam. DSCAM Ig1-Ig9 was found to be a dimer in solution and domains Ig1-Ig4 did not seem to participate in the dimerization interface. EM studies with negative staining revealed, for the first time, the overall shape of human DSCAM nine Ig domains (Ig1-Ig9) in a monomeric state; out of these, the first four *N*-terminal domains (Ig1-Ig4) were engaged in a conserved rigid horseshoe arrangement. In the presence of netrin-1, DSCAM Ig1-Ig9 was predominantly depicted as a monomer by EM (negative staining); however, the dimeric population observed was probably induced by netrin-1 with the identity of the second molecule not being clearly resolved. SAXS analysis indicated that a plausible complex between DSCAM Ig1-Ig9

and netrin-1 could be formed under the conditions used in the study, but further investigation is required to confirm complex formation.

Diffraction crystals were successfully grown for all three DSCAM constructs. DSCAM Ig1-Ig9 crystals diffracted to 9 Å resolution. Preliminary characterization of DSCAM Ig1-Ig8 at 6.7 Å resolution, showed that it crystallized in the C-centered orthorhombic lattice. For DSCAM Ig1-Ig4, two complete diffraction data sets were collected to 2.7 Å (space group C222<sub>1</sub>) and to 3.1 Å resolution (space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>), indicating that the more rigid configuration adopted by this shorter form, was more prone to crystallize and diffract X-rays to higher resolution, compared to the larger constructs. Molecular replacement did not provide a solution when Dscam Ig1-Ig4 was employed as a search model and derivatization of the crystals using heavy atoms, resulted in poor anomalous signal, not sufficient for phase determination. An additional data set was recently collected at 2.35 Å resolution from crystals belonging to space group C222<sub>1</sub>. The structure of Neurofascin, bearing a horseshoe domain organization, was used as a search model and the 3D structure of DSCAM Ig1-Ig4 was determined. The domains of DSCAM Ig1-Ig4 crystal structure adopted a horseshoe shape arrangement (as indicated by EM negative staining for DSCAM Ig1-Ig9), forming a dimer in the asymmetric unit. Dimer formation was shown to be induced by crystal packing interactions and the biological assembly of DSCAM Ig1-Ig4 seemed to be monomeric, as indicated by the results obtained for this construct in solution.

The present study demonstrates that the underlying mechanism of human DSCAM homophilic interactions differs from Dscam in *D. melanogaster*; however, this finding requires further investigation, including the 3D structure of DSCAM Ig1-Ig9 in order to elucidate its functional role. Overall, the results obtained contribute to the preliminary characterization of human DSCAM and reveal the structural differences of the two species. In addition, the applied integrated structural biology approaches determined the workflow towards understanding the structure-function relationship that dictates the interactions of DSCAM and netrin-1. Deciphering the role of cell receptors in axon guidance in vertebrates requires holistic approaches to dissect the complexity of this biological system.



## ZUSAMMENFASSUNG

Während der neuronalen Entwicklung ist die Etablierung präziser Verbindungen zwischen Neuronen entscheidend für die korrekte Ausbildung des neuronalen Netzwerks. Die Effektivität des neuronalen Netzwerks wird durch repulsive Signale gesichert, die durch homophile Interaktionen zwischen Schwester-Dendriten entstehen und somit die Ausbildung von Synapsen verhindern (Selbst-Vermeidung). Zusätzlich wird die Orientierung und das Wachstum von Axonen von Rezeptoren auf der Zelloberfläche bestimmt, die in Abhängigkeit von bestimmten Signalmolekülen (Führungssignale) anziehende oder abstoßende Reaktionen auslösen.

Das ‚Down Syndrome Cell Adhesion Molecule‘ (DSCAM), ist ein Zelloberflächen-Rezeptor, dessen Ektodomäne aus sechs Fibronectin- (FNII) und 10 Ig-ähnlichen Immunoglobulin-Domänen besteht. In Säugetieren wird er hauptsächlich in Axonen der Kommissurenzellen exprimiert und kontrolliert die neuronale Antwort durch heterophile Bindung an Führungssignale, wie z.B. Netrin-1. Ihre Interaktion löst Signale im Zytoplasma aus, die zu Chemoattraktion zwischen Axonen führen und deren Wachstum stimulieren. DSCAM vermittelt auch homophile Interaktionen, die essentiell für Zelladhäsion und -aggregation sind. Das homologe Protein aus *D. melanogaster* (Dscam) hat eine andere Funktion und scheint, ausgelöst durch homophile Bindungen, hauptsächlich in der neuronalen Selbst-Vermeidung und dendritischen Unterscheidung involviert zu sein. Das Ziel dieser Arbeit war es, die Domänen des humanen DSCAM zu identifizieren, die an der Homodimerisierung beteiligt sind, sowie heterophile Interaktionen mit Netrin-1 zu untersuchen, um die mechanistischen Unterschiede zwischen beiden Spezies zu untersuchen. Dazu wurden biophysikalische und strukturelle Studien durchgeführt, im speziellen SAXS, Elektronenmikroskopie (EM) und Proteinkristallographie.

Um die Homodimerisierung zu untersuchen, wurden drei Konstrukte des humanen DSCAM mit den N-terminalen Domänen Ig1-Ig4, Ig1-Ig8 sowie Ig1-Ig9 erstellt, exprimiert und aufgereinigt. Die Resultate der SAXS und biophysikalischen Analysen legen nahe, dass die Homodimerisierung von DSCAM und Dscam unterschiedlichen Mechanismen folgen. DSCAM Ig1-Ig9 liegt als Dimer in Lösung vor, jedoch scheinen die Domänen Ig1-Ig4 nicht an der Dimerisierung beteiligt zu sein. Mittels EM-Studien (Negativfärbung) konnte zum ersten Mal die vollständige Form von neun Ig-Domänen des humanen DSCAM (Ig1-Ig9) im monomeren Zustand gezeigt werden; von denen die ersten vier N-terminalen Domänen (Ig1-Ig4) ein

konserviertes starres Hufeisen-Arrangement bilden. In Anwesenheit von Netrin-1 zeigten die Negativfärbungs-EM-Studien DSCAM Ig1-Ig9 vorrangig als Monomer. Es ist jedoch zu vermuten, dass die beobachteten Ig1-Ig9-Dimere durch Netrin-1 induziert wurden, wobei die Identität des zweiten Moleküls nicht geklärt werden konnte. Die SAXS-Analysen weisen darauf hin, dass DSCAM Ig1-Ig9 mit Netrin-1 einen Komplex unter den gewählten Bedingungen bilden. Diese Ergebnisse müssen jedoch durch weitere Untersuchungen bestätigt werden.

Von allen drei DSCAM-Konstrukten wurden erfolgreich Proteinkristalle erzeugt. Kristalle von DSCAM Ig1-Ig9 hatten eine Auflösung von 9 Å, DSCAM Ig1-Ig8 kristallisierte in einer C-zentrischen orthorhombischen Raumgruppe und hatte eine Auflösung von 6.7 Å. Von DSCAM Ig1-Ig4-Kristallen konnten zwei komplette Datensätze gesammelt werden, mit einer Auflösung von 2.7 Å (Raumgruppe C222<sub>1</sub>) und 3.1 Å (Raumgruppe P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>). Die starrere Konfiguration des kürzeren Ig1-Ig4 scheint im Vergleich zu den langen Konstrukten dessen Kristallisation begünstigt zu haben. Die Struktur konnte jedoch weder mit ‚molecular replacement‘ unter Nutzung der bekannten Dscam Ig1-Ig4-Struktur gelöst werden noch durch Derivatisierung mit Schwermetallen, da das anomale Signal zu schwach für die Phasen-Bestimmung war. Kürzlich wurde ein weiterer Datensatz mit einer Auflösung von 2.35 Å (Raumgruppe C222<sub>1</sub>) erhalten. Unter Nutzung einer Neurofascin-Struktur, deren Domänen eine Hufeisen-Organisation aufweisen, konnte die Ig1-Ig4-Struktur gelöst werden. Sie zeigt, dass DSCAM Ig1-Ig4 eine Hufeisen-Organisation aufweist, wie es die EM-Studien mit DSCAM Ig1-Ig9 vermuten ließen, und Dimere in der asymmetrischen Einheit bildet. Das Vorhandensein der Dimere ist auf die Kristallpackung zurückzuführen, während der biologische Aufbau von Ig1-Ig4 monomerisch ist, wie die Untersuchungen des Konstrukts in Lösung vermuten ließen.

In der vorliegenden Arbeit konnte gezeigt werden, dass humanes DSCAM einen unterschiedlichen Mechanismus in der homophilen Interaktion im Vergleich zu Dscam aus *D. melanogaster* aufweist. Diese Ergebnisse erfordern jedoch durch weitere Untersuchungen, einschließlich einer 3D-Struktur von DSCAM Ig1-Ig9, um seine Funktion zu entschlüsseln. Die erhaltenen Resultate leisten einen wesentlichen Beitrag für die Charakterisierung des humanen DSCAM und decken strukturelle Unterschiede zu Dscam aus *D. melanogaster* auf. Zusätzlich beeinflussten die genutzten strukturbioologischen Anwendungen den Arbeitsablauf in Richtung eines besseren Verständnis der Struktur-Funktion-Beziehung, die Interaktion von DSCAM

und Netrin-1 bestimmen. Um die Rolle von Zellrezeptoren in der Orientierung von Axonen in Wirbeltieren zu entschlüsseln und diese komplexen biologischen Systeme zu analysieren, ist eine ganzheitliche Herangehensweise erforderlich.



# Chapter 1

## INTRODUCTION

### 1.1 Physiology of the neuron

At the end of the 19<sup>th</sup> century, Santiago Ramón y Cajal and Camillo Golgi made the first attempts to describe the components comprising the brain tissue. Golgi believed that the brain was composed of a continuous connected tissue, while Cajal suggested that the brain consisted of distinct cells. Cajal showed that the brain tissue consisted indeed of individual neuron cells, using silver impregnation, a staining technique originally developed by Golgi. His discovery confirmed his hypothesis; however, the use of any other staining technique would have been misleading, since the remaining cells would have been visualized as tangled tissue (Levitan and Kaczmarek, 2002).

Today, it is known that brain tissue is composed of two major cell types, glial and neuron cells, forming a complex nervous system. Glial cells can be categorized into microglia and macroglia (Eglitis and Mezey, 1997). Microglia are the primary immune cells in the central nervous system (CNS) and are mainly responsible for providing protection from pathogens (Prinz and Priller, 2014). Macroglia, such as astrocytes and oligodendrocyte cells in the CNS and in Schwann, enteric and satellite cells in the peripheral nervous system (PNS) (Jessen, 2004) have multiple roles in the nervous system. Among others, they provide insulation to neurons (Kettenmann and Verkhratsky, 2008), they maintain the necessary ion levels on the extracellular environment (Jessen, 2004) and they also support the neuron cells by maintaining their structure and regulating synaptic connectivity without their direct involvement in electrical signaling (Purves et al., 2001; Eroglu and Barres, 2010).

Neurons are mainly responsible for signal transduction, either intracellularly or among different cells (Purves et al., 2001). The human brain contains approximately  $10^{11}$  neurons and each neuron is connected on average with a thousand other neurons following a precise wiring path (Ranscht, 2000; Alberts et al., 2002; Bashaw and Klein, 2010). Neuronal communication in the adult nervous system is a process that is highly regulated by gene expression during neuronal development and is fundamental for the proper function of the neuronal circuit (Chédotal and Richards, 2010).

A typical neuron cell consists of an axon and dendrites. The axon is a thin structure arising from the main cell body and it can be elongated from micrometers to meters. Signal transmission is mediated with the aid of highly specific proteins located in the axonal plasma membrane that can transfer electrical signals along its length. Dendrites have a highly branched structure, shorter than the axon and they mainly arise from the cell body or originate from certain regions in the axons (e.g. in invertebrates). They are composed of small protrusions, the dendritic spines, which allow electrical signal transmission within the neuron cell body. In some cases dendrites are also involved in signal transmission with neighboring neuron cells (Alberts et al., 2002) (Figure I-1).

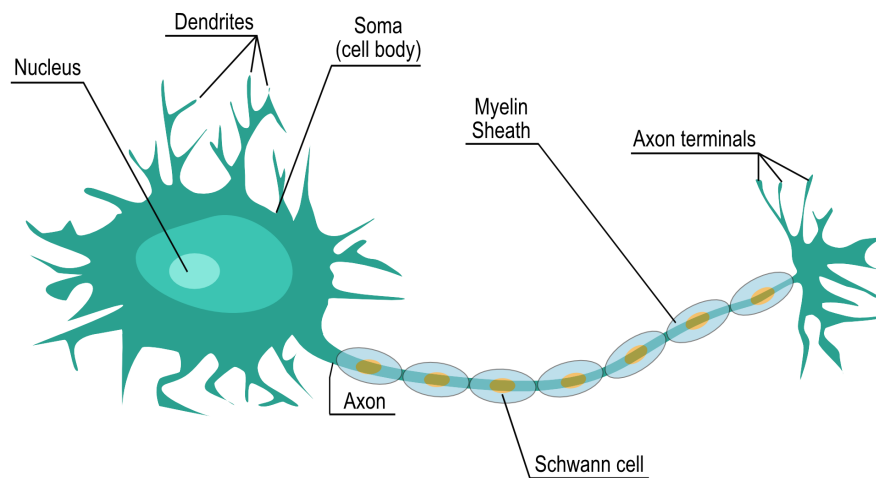
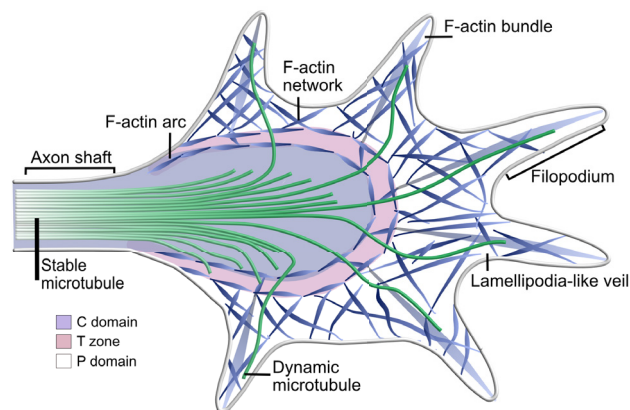


Figure I-1. Schematic representation of a typical vertebrate neuron cell.

## 1.2 Neuronal wiring

During brain development, axons follow specific pathways to establish precise neuron connections with their presynaptic targets (Chao et al., 2009; Kolodkin and Tessier-Lavigne, 2011). The specificity required from the neurons for locating their binding partners, is dictated by a variety of mechanisms, which among others involve diffusible guidance molecules (chemotropic cues) that influence the axonal growth and synaptogenesis, adhesive molecules that activate intracellular signaling and cell receptors that either promote or prohibit the axonal growth (Dickson, 2002). In particular, guidance cues exhibit high diversity, allowing axons to modify their responses depending on the receptors present on the neuron surface. This process influences axonal navigation and along with second messenger signaling pathways

that mediate dynamic cytoskeletal reorganization, lead to the final axon movement (Chauvet and Rougon, 2009; Bashaw and Klein, 2010; Dent et al., 2011). Cell surface receptors are located at the tip of the axon, the growth cone, which facilitates axonal movement and growth during development. The growth cone is an actin-based motile structure, rich in microtubules that maintain the shape of the neuron and provide structural support (Lowery and van Vactor, 2009). More specifically, it is mainly driven by subsequent polymerization and depolymerization of the actin filaments, which along with motor proteins (such as myosin II) (Medeiros et al., 2006) promote the movement of the neuron, while neurofilaments (intermediate filaments) maintain the neuron structure (Alberts et al., 2002; Burnette et al., 2008; Levitan and Kaczmarek, 2002). The projections of the growth cone are called filopodia or microspikes (finger-like extensions), which are surrounded by a membrane composed by mesh-like F-actin network, the lamellipodia-like veils (flat sheet-like extensions). Both filopodia and lamellipodia are constantly in motion, enhancing the final axonal elongation (Alberts et al., 2002; Dent et al., 2011) (Figure I-2).

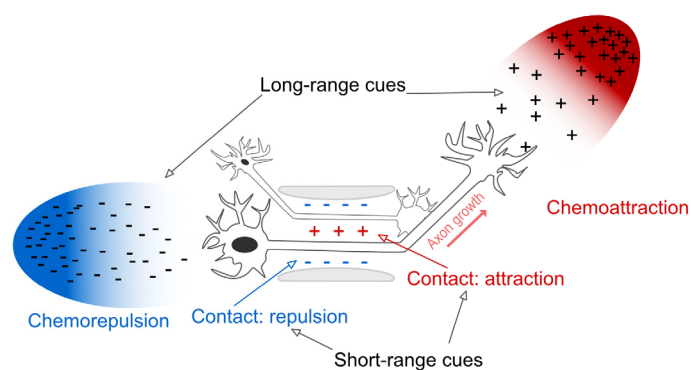


**Figure I-2.** Schematic representation of the growth cone which is divided into three areas; the Peripheral (P) and the Central (C) domains, along with the Transition zone (T). The P domain contains individual dynamic microtubules and long actin filaments, the F-actin bundles, which compose the filopodia. The latter along with the lamellipodia create a highly dynamic P domain in which cell receptors and adhesion molecules are mainly located. The C domain exhibits less dynamic motion and contains organelles, vesicles and stable microtubules, which enter the growth cone arising from the axon shaft, maintaining the normal growth cone morphology. The T zone is located between the P and the C domains and contains mainly motor proteins and F-actin arcs (actomyosin contractile structures). The arc structures are composed of antiparallel bundles of actin that interact with microtubules, transporting them to the central domain. Image adapted from (Lowery and van Vactor, 2009).

Initially, a hypothesis that neurons carry specific molecules, which guide the axons to certain directions, was investigated by Langley (Langley, 1895) and later on by

Sperry (Sperry, 1963), who performed neurological experiments on cats and amphibians, respectively. Their studies involved dissecting the optic nerves of their animal models, which resulted in neuron regeneration and formation of functional connections. Based on these observations, Sperry originally suggested that axons locate their targets and establish neuron connections, with the assistance of molecules (molecular tags) present on the neuron surface of the axon and their counterparts located on the target surface. Therefore, neurons must be chemically specified with certain molecules in order to promote axonal growth (Chemoaffinity hypothesis) (Sperry, 1963). His initial hypothesis, was later modified stating that guidance cues, are also essential for allowing a correct neuron targeting (Meyer, 1998).

During the last twenty years, a plethora of experimental data and the discovery of guidance cues, have supported the chemoaffinity hypothesis. More specifically, a variety of molecules with multifunctional properties that have been identified as guidance cues, along with cell surface receptors, provide a key role in neuron recognition, specificity and axonal growth. Guidance cues may act either at short range, by being present on the growth cone, or at longer range, secreted from a distant target (Kolodkin and Tessier-Lavigne, 2011). They mediate attraction or repulsion depending on the surface cell receptors they respond to. Thus, in the case of chemoattraction, the growth cone turns towards the signal source, resulting in axonal extension, whereas in the case of chemorepulsion, the axon moves away from the source (de Castro et al., 1999) (Figure I-3).



**Figure I-3.** The diversity of axon guidance. Guidance cues can act either at short or long range mediating attraction or repulsion responding to different receptors on the neuron cells. Image adapted from (Kolodkin and Tessier-Lavigne, 2011).



Studies have shown that certain proteins are responsible for allocating different specificity to the axons that act as diffusible chemotropic cues, such as netrins (Kennedy et al., 1994; Serafini et al., 1996), semaphorins (Kolodkin et al., 1993), slits (Brose et al., 1999) and Ephrins (Tessier-Lavigne, 1995) along with a broad range of receptors e.g. Down Syndrome Cell Adhesion Molecule (DSCAM) (Andrews et al., 2008), deleted in colorectal cancer (DCC) (Keino-Masu et al., 1996), roundabout (Robo) (Kidd et al., 1999), uncoordinated5 (UNC-5) (Leung-Hagesteijn et al., 1992), plexins (Tamagnone et al., 1999) and Eph (Klein, 2004) (Table I-1). Additionally, the growth cone contains molecules that are involved both in physical cell adhesion and synapse formation and also in neuronal wiring during development, like cell adhesion molecules (CAM), members of the Immunoglobulin superfamily (Ig), fibronectins (FN), laminins and cadherins (Tomaselli et al., 1988; Kolodkin and Tessier-Lavigne, 2011) (Table I-1).

Table I-1. Cell adhesion and axon guidance molecules with their receptors.

Adhesion factor	Receptor	Guidance molecule	Receptor
Fibronectin	Integrin	Semaphorin	Neuropilin
Laminin			Plexin
CAM	CAM	Netrin/UNC-6	DCC (vertebrates)/UNC-40
DSCAM	DSCAM	( <i>Caenorhabditis elegans</i> )	( <i>C. elegans</i> )
			Neogenin
			DSCAM
			UNC-5 (vertebrates)
Cadherin	Cadherin	Ephrin	EphA & EphB
		Slit	Robo

### 1.3 Axon guidance for neuronal wiring

During embryonic development in vertebrates, the correct wiring of the spinal cord through axon guidance is essential in order to neurally coordinate the two sides of the body. The floor plate located at the ventral midline of the spinal cord, is composed of epithelial cells and is considered as an intermediate target that influences axon guidance (Odenthal et al., 2000; Chao et al., 2009). More specifically, commissural axons, which connect the CNS across the spinal cord, differentiate in the dorsal area of the latter and follow a pathway towards the ventral midline of the cord, by crossing the floor plate (Colamarino and Tessier-Lavigne, 1995). A key feature of the commissural axons is that they can switch their response to specific guidance cues, by altering their sensitivity as they travel across the floor plate. This is

achieved by changing the type of receptors expressed on their surface (Shirasaki et al., 1998). In that way, the system ensures that the axons, which have initially responded to a specific guidance cue, will lose sensitivity for it while remaining in their target region (Figure I-4). The repulsion signal, generated by the floor plate, is sensed by the axons and prevents them from returning to their original position in the spinal cord, ensuring a precise neuronal wiring (Alberts et al., 2002). Studies in zebrafish (Bernhardt et al., 1992) and mice mutants lacking a floor plate (Kadison et al., 2006), have shown that its role is substituted by other cells acting as intermediate targets. In invertebrates, like in *Drosophila melanogaster*, the floor plate is substituted by specific midline glial cells that promote axon guidance (Klämbt et al., 1991).

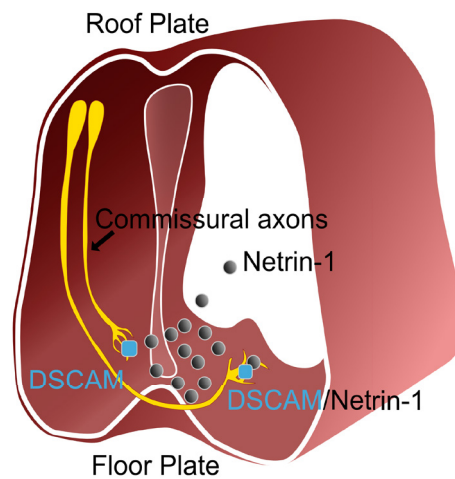


Figure I-4. Schematic representation of the neural tube in the developing spinal cord in vertebrates. The guidance cue netrin-1 (shown in black), which is secreted by the cells of the floor plate, attracts the commissural axons (shown in yellow) to the floor plate, upon DSCAM (shown in blue) expression. The axons expressing DSCAM can mediate turning responses to netrin-1 (chemoattraction) (Ly et al., 2008).

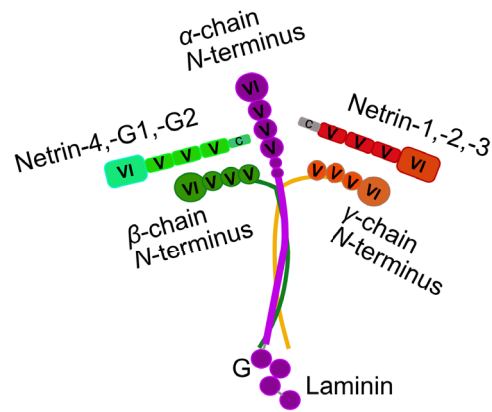
Signaling factors are secreted either by the floor plate, the roof plate (located in the dorsal midline of the cord), or by the ventral two-thirds of the spinal cord, and guide specific receptors of the commissural axons towards the spinal cord (Kennedy et al., 1994; Chizhikov and Millen, 2004). This activation leads to membrane depolarization, resulting in an intracellular  $\text{Ca}^{2+}$  release. The change in  $\text{Ca}^{2+}$  levels activates the extension mechanism and the actin reorganization induces dynamic changes in the cytoskeleton, which lead to movement of the growth cone towards the gradient source (Hong et al., 2000; Bashaw and Klein, 2010). Axons that do not express the corresponding receptors are not accepted and they are targeted back to the roof plate (Alberts et al., 2002). On the contrary, chemorepulsion mediates

membrane hyperpolarization. For example, in the case of semaphorins, a local change in the calcium influx, results in the activation of the plexin receptor, which in turn activates certain ion channels in the membrane. Therefore, calcium levels guide the axon away from its source, mediating repulsion (Nishiyama et al., 2008).

#### 1.4 Netrin as a guidance cue

Netrins are secreted proteins that had been originally characterized as chemoattractants, however, both *in vivo* and *in vitro* studies, demonstrated that they also promote chemorepulsion (Colamarino and Tessier-Lavigne, 1995) and their role as bifunctional guidance cues has been evolutionary conserved (Kolodkin and Tessier-Lavigne, 2011). They are found both in vertebrates and invertebrates with UNC-6 being the first netrin identified in *C. elegans* (Hedgecock et al., 1990; Ishii et al., 1992). There are two netrins in *D. melanogaster*, netrin-A and netrin-B (Harris et al., 1996), four proteins in mammals netrin-1, netrin-3, netrin-4, netrin-5 (Yamagishi et al., 2015) with netrin-2 being recently discovered in *Gallus gallus*. There are also two membrane associated glycosylphosphatidylinositol (GPI) linked proteins in mammals netrin-G1 and netrin-G2 (Kolodkin and Tessier-Lavigne, 2011). Netrins guide many different axons by acting either over a short (Brankatschk and Dickson, 2006; Rajasekharan and Kennedy, 2009) or over a long distance (Kennedy et al., 1994) and they mediate repulsion or attraction depending on the receptor expressed on the neuron surface (Kolodkin and Tessier-Lavigne, 2011). Within the spinal cord of the vertebrates, two netrins exist. Netrin-1, which is only present in the floor plate and netrin-2, mainly in the ventral half of the spinal cord present in *Gallus gallus domesticus* (Keino-Masu et al., 1996; Kennedy et al., 1994).

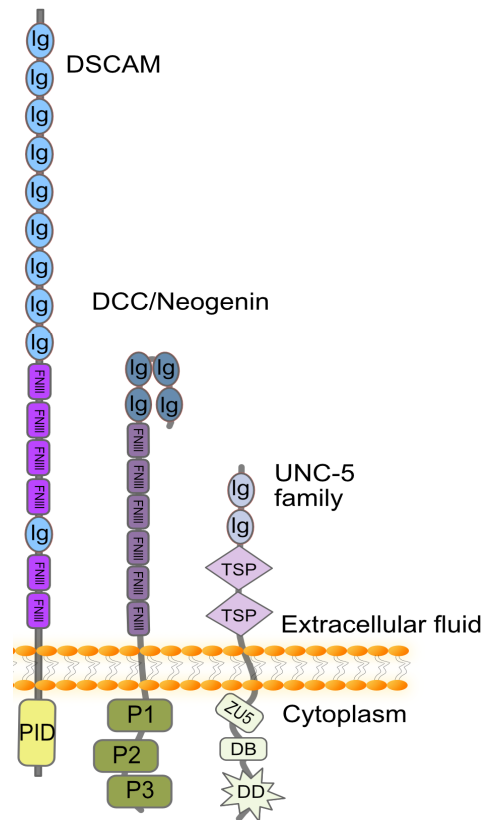
Netrin-1 is expressed in the developing neuron system, influencing axon guidance and synaptogenesis. In the non-neural systems it is highly expressed in tissues, such as heart (Zhang and Cai, 2010), lungs (Liu et al., 2004) pancreas (Yebra et al., 2003) and the visual system (Sugimoto et al., 2001), where netrin-1 regulates cell migration, cell adhesion and survival (Rajasekharan and Kennedy, 2009). It is composed of a laminin domain VI, which is globular, followed by three epidermal growth factors EGF-like repeats of the  $\gamma$ -chain of laminins V (EGFV-1, EGFV-2, EGFV-3). These N-terminal domains are linked to a C-terminal netrin-like module (Lai Wing Sun et al., 2011) (Figure I-5).



**Figure I-5.** Schematic representation of netrin family. Netrins are members of the laminin superfamily (in purple) and consist of three major domains, V, VI and C. The *N*-terminal laminin encodes domains VI and V (in red), which are homologous to the corresponding laminin domains VI, V of the  $\gamma$ -chain (in orange). Similarly, domains VI, V (in light green) which correspond to Netrin-4, G1 and G2, are homologous to the  $\beta$ -chain of laminins (in dark green). Image adapted from (Ko et al., 2012).

#### 1.4.1 Cell receptors of netrin-1

Netrin-1 is associated with a significant number of cell surface receptors for the activation of chemotropic responses and adhesive mechanisms. In mammals, several receptors that mediate netrin-dependent attraction, have been identified, such as DCC (Keino-Masu et al., 1996; Chen et al., 2013; Finci et al., 2014;), neogenin (a DCC paralogue in vertebrates) (Bell et al., 2013; Xu et al., 2014) and DSCAM (Ly et al., 2008; Liu et al., 2009). DCC orthologues are present in *C. elegans* (UNC-40) (Chan et al., 1995) and in *D. Melanogaster* (Frazzled) (Kolodziej et al., 1996) promoting chemoattraction as well. On the contrary, axonal chemorepulsion in vertebrates, is mediated when UNC-5(a-d) (also found in *D. melanogaster* and in *C. elegans*) is present on the neuron surface (Hong et al., 1999; Keleman and Dickson, 2001) (Figure I-6).



**Figure I-6.** Schematic representation of selected cell surface receptors interacting with netrin-1, such as DSCAM, DCC with its paralogue in vertebrates: neogenin and homologue members of DCC: the UNC-5 family. Ig: Immunoglobulin domain, FNIII: Fibronectin type III domain, PID: Pak-interacting domain, P1, P2, P3: phosphoserine (P1, P3) and phosphothreonine (P2) conserved sequence domains, TSP: thrombospondin type 1 domain, ZU5: zona occludens 5 domain, DB: binding domain of DCC, DD: death domain. Image adapted from (Lai Wing Sun et al., 2011).

## 1.5 Deleted in Colorectal Cancer

The gene of DCC was originally characterized as a potential tumor suppressor, detected on the human chromosome 18q21, which is usually absent in colorectal carcinomas (Fearon et al., 1990). DCC is involved in mediating the transition from proliferation to terminal differentiation and its loss is implicated in the majority of colorectal and other types of cancer and probably also leads to tumor metastasis (Krimpenfort et al., 2012). DCC binds to netrin-1 inducing a signaling complex. Upon the absence of netrin, the complex dissociates leading to cell apoptosis. In cancer cases, the tumor cells switch off this apoptotic pathway, by deleting the cell receptor of DCC, resulting in selective cellular survival (Mazelin et al., 2004). In vertebrates during development, DCC is present on the growth cones of spinal commissural

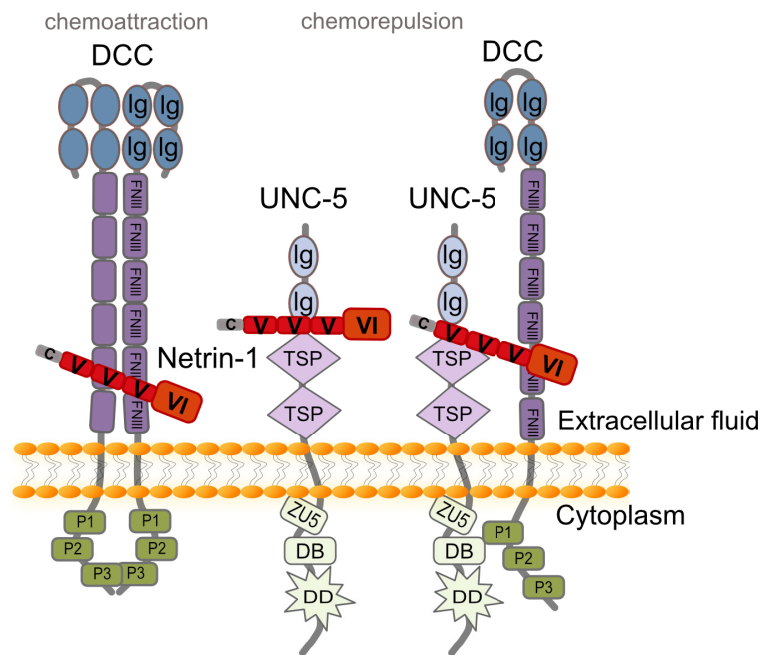
neurons and it is required for promoting the axonal growth in response to netrin-1. Loss of either netrin-1 or DCC results in loss or mistargeting of several commissural axons *in vivo* (Keino-Masu et al., 1996).

The DCC protein family consists of four Ig domains, arranged in a horseshoe conformation (Chen et al., 2013) and six FNIII repeats in its extracellular region, followed by a transmembrane single domain. The intracellular domain comprises three sequence conserved phospho-domains (P1, P2, P3), whose role is to accumulate cytoplasmic signal transduction molecules (Rajasekharan and Kennedy, 2009) (Figure I-6). Studies have shown that the extracellular domains of DCC are required for binding to netrin-1, which induces the formation of a homodimer, involving also the P3 intracellular domain. This binding promotes chemoattraction, which leads to the reorganization of the cytoskeleton, with the activation of several proteins within the cell (e.g. Rho family of GTPases, Src-kinases, Ser/Thr kinases) (Rajasekharan and Kennedy, 2009; Lai Wing Sun et al., 2011). More specifically, it has been proposed that DCC binds to netrin-1 through the extracellular N-terminal FNIII domains FN4 and FN5 (Geisbrecht et al., 2003; Kruger et al., 2004), a view that was also supported by structural studies (Xu et al., 2014).

Additional structural studies between netrin-1 and domains FN5-FN6 of DCC, revealed homodimerization of DCC upon binding to netrin-1, providing evidence that these FN domains are also essential for inducing binding (Finci et al., 2014). The 3D structure revealed that the binding sites of DCC were located in two different regions of the V domain of netrin-1, underlying the important role of domain FN5. One binding site (BS1) was located at the tip of the EGFV-3 domain of netrin, and the amino acids implicated in the interactions formed, proved to be evolutionary conserved. Binding at this site involved only domain FN5, in accordance with the structural findings obtained by Xu and co-workers for the specific domain (Xu et al., 2014). The second binding site (BS2) involved both domains FN5 and FN6, which interacted with netrin-1 domain EGFV-1 (Figure I-7). Based on neuronal experiments with netrin-1, UNC-5 and DCC mutants, it was proposed that BS1 was specific for DCC binding, whereas BS2 could accommodate other receptors such as UNC-5, offering distinct signaling (Finci et al., 2014).

These findings provide an example of how the axons switch their response from attraction to repulsion when distinct receptors are present on the cell surface, e.g. in the case of netrin-1 and its counterparts, the complex of DCC:netrin-1 mediates chemoattraction, whereas UNC-5:netrin-1 induces chemorepulsion. In vertebrates, it

has been shown that many neurons require both UNC-5 and DCC on their cell surface for repulsion that occurs upon netrin-1 binding to their extracellular domains (Colavita and Culotti, 1998; Hong et al., 1999). This binding mode potentially triggers the formation of a UNC-5:DCC complex, involving also interactions of their cytoplasmic domains, necessary for the ternary structure (Finci et al., 2014; Lai Wing Sun et al., 2011) (Figure I-7).



**Figure I-7.** Schematic representation showing the biofunctional role of guidance cues in axon guidance. Binding of netrin-1 (shown in red) to DCC extracellular domains induces chemoattraction through clustering of DCC molecules with the involvement of cytoplasmic domains P3, whereas upon binding to UNC-5 induces chemorepulsion, in cases that short-distance repulsion is required. When netrin-1 binds simultaneously to both UNC-5 and DCC, chemorepulsion is also mediated (in long range repulsion cases), inducing cytoskeletal changes through binding of P1 and DB domains. This is an example of how an axon can switch its response, depending on the receptors that are expressed on its surface. Image adapted from (Cirulli and Yebra, 2007).

## 1.6 Down Syndrome Cell Adhesion Molecule

Cell adhesion is an essential process, found in several tissues, for maintaining cellular structure. This molecular regulation of morphogenesis is based on the differential gene expression of cell adhesion molecules (CAM). When this mechanism is triggered, a number of surface modifications occur, which provide the cells with specific motilities. This synchronization of the CAM expression helps decimating the non-acceptable connections between the cells (Edelman, 1984). More

specifically, CAM and members of the Ig superfamily (e.g. DSCAM), provide the bridge for cell-cell communication, through homophilic and heterophilic interactions through mechanical processes (Dalva et al., 2007). Besides that mechanism, it has been additionally proposed that they are also involved in axon guidance and in signaling pathways regulation, during neuronal development (Kolodkin and Tessier-Lavigne, 2011). Initially, it was suggested that only CAM guide axons through cell adhesion (Harrelson and Goodman, 1988; Lin and Goodman, 1994), but recent studies showed that axon guidance and outgrowth is also mediated by responding to signals arising from other axons mainly through heterophilic interactions (Kolodkin and Tessier-Lavigne, 2011).

The Ig superfamily of DSCAM is a cell surface receptor and its gene was initially identified in chromosome band 21q22, which is responsible for many of the phenotypes in Down Syndrome (trisomy 21) (Yamakawa et al., 1998). It was proposed that the presence of three copies of DSCAM, observed in Down Syndrome (DS), might affect cell adhesion by altering cell properties, resulting in neuron clustering. In addition, postmortem studies in DS cases have revealed alterations in axons and in dendrite spine morphology, proposing that DSCAM plays an important role in axon guidance (Antonarakis and Epstein, 2006). In humans, the second gene (DSCAML1) is located in chromosome 11q23 and is related to Tourette and Jacobsen Syndromes (Agarwala et al., 2001).

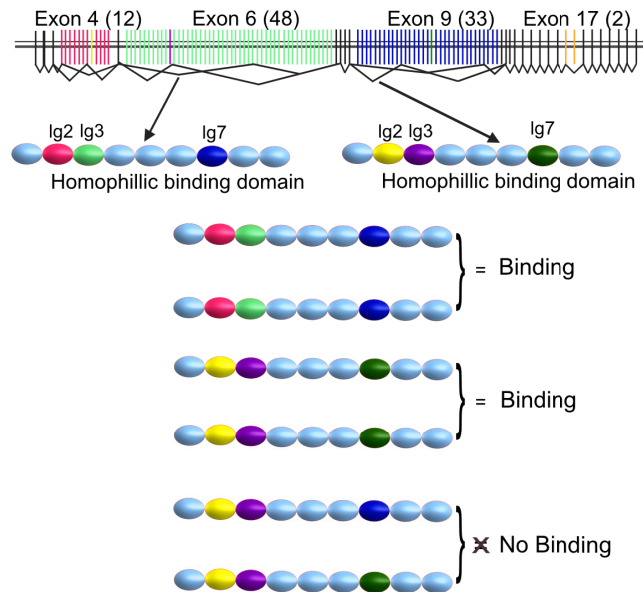
### **1.6.1 Dscam in *Drosophila melanogaster***

Previous studies have shown that Dscam, a *D. melanogaster* homologue gene of human DSCAM, encodes through alternative splicing, more than 38,000 variable isoforms and is involved in the formation of the neuronal circuit (Schmucker et al., 2000). The extracellular domains of Dscam are composed of ten Ig-like C2-set domains and six FNIII domains. The Ig-like domains contain 70-110 amino acids and they share a characteristic Ig-fold with a Greek-key-sandwich-core structure formed by two  $\beta$ -sheets of antiparallel  $\beta$ -strands.

More specifically, the Dscam gene in *Drosophila* contains three arranged groups of alternative exons, which encode 12, 48 and 33 isoforms for the domains Ig2, Ig3 and Ig7, respectively. The locus also encodes two variations of the transmembrane domains and alternative splicing produces also four different C-terminal tails. Due to the fact that mRNA splicing in each group of exons is independent of the others, it gives the possibility for generation of a total of 19,000 different extracellular domains



in 38,000 variations (Schmucker et al., 2000; Zipursky and Grueber, 2013) (Figure I-8).

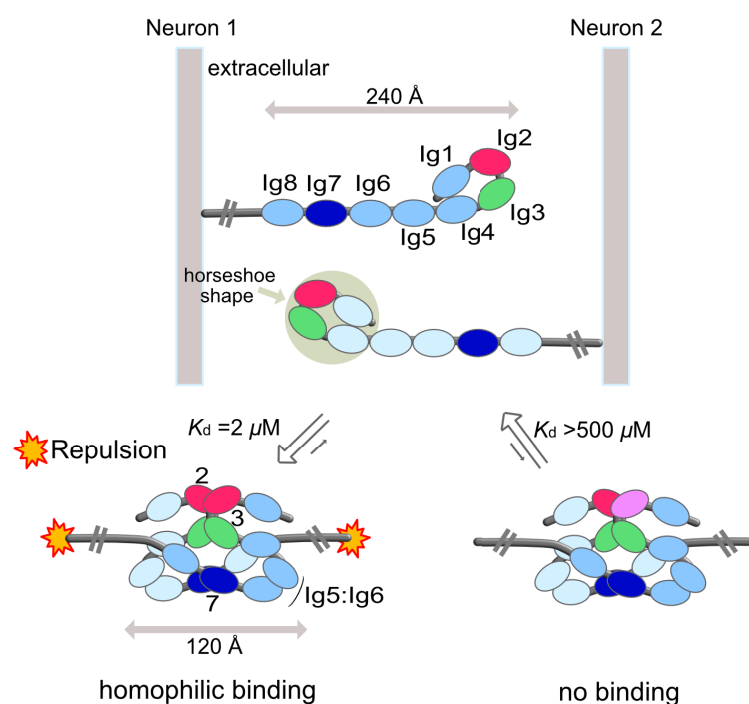


**Figure I-8.** Alternative m-RNA splicing of Dscam gene in *D. melanogaster* that generates multiple isoforms depicting the molecular variability of Dscam protein. The gene consists of four blocks of exons, each responsible for encoding alternative variants of Ig domains. Exon 4, 6 and 9 encode Ig2, Ig3 and Ig7, respectively and the combination of different exons results in 19,008 ectodomains. For the transmembrane domain, exon 17 consists of two variants encoding this region of the protein. For Dscam homophilic binding all three domains (Ig2, Ig3, Ig7) should match. In case there is a mismatch (even one is sufficient), binding does not occur. Image adapted from (Zipursky and Grueber, 2013).

Several biochemical and cell aggregation assays revealed the role of the domains variability. Experiments, where more than 3,000 different isoforms were tested, demonstrated that homophilic binding (with a few exceptions) occurred only when domains Ig2, Ig3 and Ig7 were matching with the corresponding ones from the other isoforms (Wojtowicz et al., 2004, 2007) (Figure I-8). A  $K_d$  estimation of the homodimers when the three domains were matching, was of 1-6  $\mu\text{M}$  (Wu et al., 2012) whereas matching only between domains Ig3 and Ig7 did not lead to dimer formation. This result underlines the specificity required for homodimerization between the different isoforms.

Structural studies of two Dscam isoforms comprising the first four Ig domains (Dscam Ig1-Ig4), confirmed the specificity required for domains Ig2 and Ig3 (Meijers et al., 2007). Both structures share a U-shape (horseshoe) arrangement involving isoform pair matching of domains Ig2 and Ig3 in an antiparallel orientation. Further studies of Dscam ectodomain consisting of eight Ig N-terminal domains (Dscam Ig1-Ig8), confirmed the aforementioned and also illustrated the role of Ig7 domain in

homophilic binding (Sawaya et al., 2008). This domain, pairs with the Ig7 of the same isoform in an antiparallel way, consistent with the pairing observed for domains Ig2 and Ig3. The 3D structure also revealed that domains Ig5 and Ig6 are essential for allowing the homophilic binding of Ig7 domain. More specifically, Ig5 and Ig6 domains are involved in intramolecular interactions, forming a 'bend' interface that stabilizes the chain turn, bringing domain Ig7 in an antiparallel axis for homophilic pairing (Sawaya et al., 2008; Schmucker and Chen, 2009) (Figure I-9). Although, the binding interface of Dscam homodimerization in *Drosophila*, contains all five domains Ig2, Ig3 and Ig5-Ig7, domains Ig2, Ig3 and Ig7 are required for specificity and recognition. All the domains are finally arranged in a double S-shape configuration, associating in a two-fold symmetric dimer, with domains Ig1-Ig4 forming a rigid horseshoe-like conformation (Figure I-9).



**Figure I-9.** Schematic representation of the proposed model for Dscam homophilic interactions based on the crystal structure of Dscam Ig1-Ig8 in *Drosophila* (PDB entry 3DMK) (Sawaya et al., 2008). Upon homodimerization, Ig domains are arranged in a double S-shape conformation. Out of these, Ig1-Ig4 are engaged in a horseshoe configuration. The specificity required for homophilic binding is attributed to domains Ig2, Ig3 and Ig7 of each monomer. Domains Ig5 and Ig6 form intramolecular interactions that allow and stabilize the chain turn, bringing domain Ig7 in an antiparallel axis for homophilic pairing. When homophilic binding occurs ( $K_d = \sim 2 \mu\text{M}$ ), a repulsion signal is generated, which is then transduced to the cytoplasmic domain of Dscam, facilitating cytoskeletal rearrangements. This event leads to self-avoidance and discrimination among sister dendrites. When the variable Ig domains do not match, binding does not occur ( $K_d > 500 \mu\text{M}$ ). The  $K_d$  values were determined by analytical ultracentrifugation (Wu et al., 2012). Image adapted from (Sawaya et al., 2008 and Zipursky and Grueber, 2013).

Dscam in *D. melanogaster* is expressed in axons and dendrites during development and is involved both in homophilic and heterophilic interactions. It is implicated in axon guidance in certain neuron types (optic nerve) (Schmucker et al., 2000) and in neuronal wiring through cell recognition and discrimination. Loss of Dscam results in dendrite self-crossing and fasciculation (Kise and Schmucker, 2013). More specifically, Dscam through homophilic interactions provides discrimination among the dendrites arising from the same neuron cell. Sister dendrites usually express the same Dscam isoforms in their cell surface and when in contact, Dscam pairs with the neighboring dendrite. Homophilic binding instantly generates a repulsion signal inducing cytoskeletal rearrangements. This results in self-discrimination among sister dendrites, preventing undesirable synapse formation (Kise and Schmucker, 2013; Matthews et al., 2007) (Figure I-9). This process is called self-avoidance and instructs a selective recognition mechanism for neuronal discrimination among neighboring neurons (non-self) and sister dendrites on the same neuron (self). The high molecular diversity of Dscam and its domain specificity, provide each neuron with a distinct identity (Neves et al., 2004; Kise and Schmucker, 2013). This is essential for neuronal recognition and discrimination to ensure correct wiring and formation of a functional neuronal system (Chen et al., 2006).

Watson and co-workers showed that Dscam was involved in heterophilic binding interactions in the immune system of insects. It was suggested that Dscam diverse isoforms attract bacterial pathogen molecules and their binding resulted in phagocytosis, demonstrating a potential role in immune signaling (Watson et al., 2005) (Figure I-10). Dscam was also found to be implicated in axon guidance, through dendritic targeting signals, arising from netrin-A, netrin-B (Andrews et al., 2008) and Frazzled (Matthews and Grueber, 2011); however, the underlying mechanism is still under investigation.

### **1.6.2 DSCAM in vertebrates**

DSCAM studies, performed mainly in mammals, have shown that it is involved in commissural axon guidance and dendritic branching during embryonic neuronal development, as well as in heterophilic interactions with other cell receptors when present, such as DCC (Liu et al., 2009; Ly et al., 2008).

DSCAM mediates axon turning in response to netrin-1 concentration gradient. More specifically, *in vivo* studies in mice demonstrated DSCAM role in commissural axon development. In the presence of DSCAM, axons managed to reach the floor

plate, whereas in knockdown experiments in mice lacking DSCAM siRNA, axons failed to cross the ventral midline (Ly et al., 2008). Additional studies with netrin-1 and DSCAM, showed that knockdown of DSCAM prevented netrin-induced axonal growth and commissural attraction. Therefore, it was concluded that DSCAM response to netrin-1 endows axon guidance and growth through chemoattractive responses.

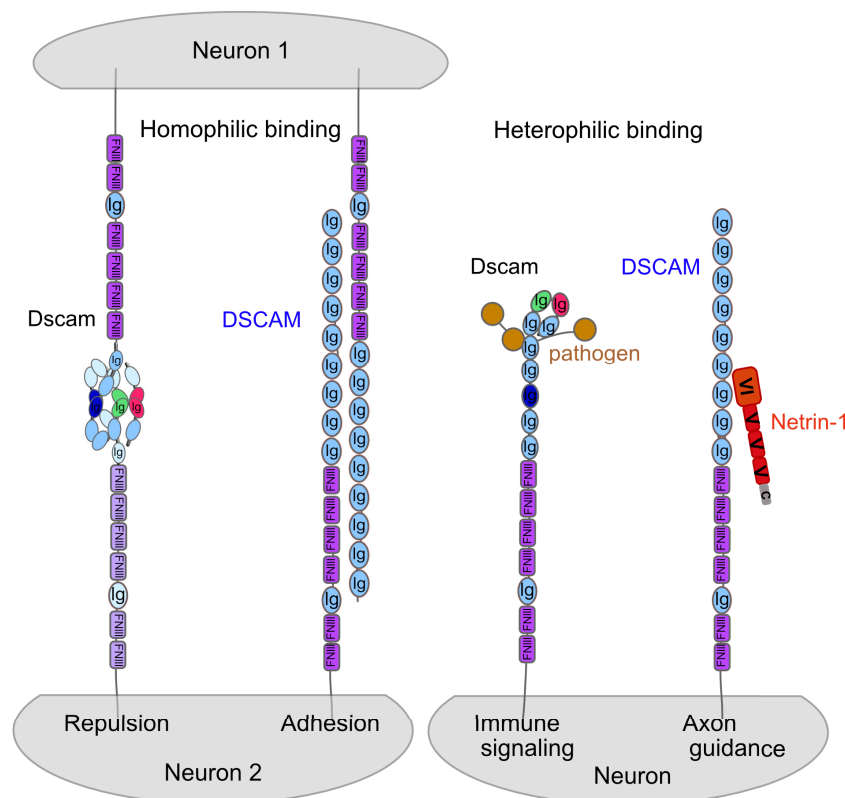
Further studies on the interactions formed upon DSCAM:netrin-1 binding, were performed with co-immunoprecipitation and activity assays. To determine domain specificity, in-solution-binding assays were performed with DSCAM domains Ig1-Ig6, Ig7-Ig9, Ig1-Ig9 and FN1-FN6. The results demonstrated that binding possibly occurred among domains Ig1-Ig9 and Ig7-Ig9, suggesting that domains Ig7-Ig9 should contribute to netrin binding (Ly et al., 2008). Additional studies for examining DSCAM oligomeric state in response to netrin-1, showed that DSCAM does not homodimerize upon netrin binding, suggesting that either DSCAM acts as a monomer, or it requires another co-receptor for signal response (Ly et al., 2008) (Figure I-10). However, the results obtained from cell binding assays, imply the presence of an additional molecule, induced by DSCAM expression, that might also bind to netrin-1 (Liu et al., 2009). Therefore, further research is required to elucidate whether another protein is also essential for DSCAM:netrin-1 complex formation.

The role of DSCAM in association with DCC was also examined both in the presence and absence of netrin-1. *In vivo* assays demonstrated that both receptors were localised on the cell surface of some commissural axons, forming a heterocomplex, via interaction of their transmembrane domains. In the presence of netrin-1, complex interactions were abolished, indicating that netrin-1 interferes with DSCAM:DCC heterodimerization. Therefore, it was proposed that either DSCAM:DCC complex is required for another signaling mechanism, or the complex is maintained until the concentration of netrin-1 reaches a critical point, at which proteins respond, by dissociating the complex previously formed (Ly et al., 2008).

Besides heterophilic interactions, DSCAM in mammals was shown to be also involved in homophilic binding (homodimerization), which is crucial for promoting cell adhesion and aggregation during synaptic targeting (Agarwala et al., 2001; Yamagata and Sanes, 2008) (Figure I-10).

Despite the absence of molecular variability (only two isoforms are expressed in humans, DSCAM and DSCAML1), in certain cell types, the role of Dscam in *Drosophila* concerning chemorepulsion is also conserved in mammalian DSCAM.

More specifically, studies performed on DSCAM in the mouse retina, concluded that DSCAM is involved in self-avoidance through homophilic interactions, to prevent neurons from fasciculation (Fuerst et al., 2008). However, this theory proposes that DSCAM in this case, is likely involved in anti-adhesive functions, rather than promoting directly repulsion in specific cell types (Fuerst and Burgess, 2009; Garrett et al., 2012). Since DSCAM in mammals lacks isoform diversity, in contrast to Dscam in *Drosophila*, it was suggested that it is only implicated in self-avoidance but not in self-discrimination, which requires the expression of a variety of different isoforms (Fuerst et al., 2008; Schmucker and Chen, 2009). Further studies are required to provide evidence on DSCAM role in self-avoidance and axon guidance in order to uncover the molecular mechanisms underlying these processes in vertebrates and understand how neuronal wiring is achieved.



**Figure I-10.** Schematic representation of DSCAM in humans (shown in blue) and Dscam in *D. melanogaster* (shown in black), which summarizes their implication in homophilic (left image) and heterophilic (right image) interactions. Dscam is engaged in homophilic interactions mediating repulsion, which leads to self-avoidance and discrimination, whereas DSCAM mediates attraction, through cell adhesion, during synaptic targeting. Anti-adhesive function is maintained only in some neuron types (e.g. mouse retina), but not self-avoidance. Upon heterophilic interactions, Dscam binding to different pathogen molecules is suggested to be important for immune signaling processes, whereas DSCAM binding to guidance cues, like netrin-1, mediates axon guidance and growth during neuron development. Image adapted from (Schmucker and Chen, 2009).



## 1.7 Aim of the study

Biochemical and structural studies performed on Dscam in *D. melanogaster*, propose a mechanism of its homodimerization. However, the corresponding mechanism underlying DSCAM homophilic and heterophilic interactions in humans is still under investigation. DSCAM lacks molecular diversity, compared to the large number of isoforms expressed in Dscam. In the frame of this thesis the hypothesis that DSCAM function is dictated by a different dimerization profile, compared to the one observed in insects, was examined. The aim of the research conducted, was to identify the interactions that drive human DSCAM homophilic binding and to assess potential differences of its homodimerization, with Dscam in *Drosophila*. The role of DSCAM in axon guidance was also explored, through DSCAM interactions in association with netrin-1, with the aim to elucidate whether DSCAM homodimerization is abolished upon binding to netrin-1 and to determine the domains implicated in heterophilic binding. To this end, structural and biophysical studies using SAXS, EM and X-ray protein crystallography were performed with DSCAM and DSCAM in the presence of netrin-1.





## Chapter 2

### MATERIALS AND METHODS

#### 2.1 Materials

The chemicals, laboratory consumables, equipment, antibodies, enzymes and primers used in the frame of the present thesis are described in the Appendix Section.

#### 2.2 Methods

This section describes in detail the procedures followed for performing the experimental work that has been included in the research conducted.

##### 2.2.1 Constructs

Down Syndrome Cell Adhesion Molecule is a member of the Ig superfamily. It consists of nine Ig-like C2-set domains located at the *N*-terminus, followed by six FNIII domains with a tenth Ig-like C2-set domain lying between the fourth and the fifth FNIII domain. The protein also consists of a transmembrane and a cytoplasmic domain at the *C*-terminus. Studies performed in the frame of this thesis, on human DSCAM (*Homo sapiens* UniProtKB-O60469), focused on the structural characterization of its extracellular domain. The DNA sequences encoding the *N*-terminal eight Ig domains of DSCAM (DSCAM Ig1-Ig8), along with the ninth domain (DSCAM Ig9), were obtained as synthetic genes designed by GenScript, optimized for the mammalian expression system. In an effort to identify the different protein-protein interactions formed, two additional constructs were designed, including the *N*-terminal four and nine Ig domains of DSCAM, DSCAM Ig1-Ig4 and DSCAM Ig1-Ig9, respectively (Table M-1).

Table M-1. List of constructs for protein expression in Human Embryonic Kidney cell line.

Protein ID	Construct	Domains	HEK 293T	HEK 293S (GnTI)
DSCAM O60469	DSCAM Ig1-Ig8 (Synthetic)	Ig1-Ig8 a.a. residues: 18-783	+	+
	DSCAM Ig9 (Synthetic)	Ig9 a.a. residues: 787-883	+	
	DSCAM Ig1-Ig9 (PCR)	Ig1-Ig9 a.a. residues: 18-883	+	+
	DSCAM Ig1-Ig4 (PCR)	Ig1-Ig4 a.a. residues: 18-401	+	+

Table M-2. Selected web tools for protein construct design and structure prediction.

Name	Function	Reference
GlyProt	<i>In-silico</i> Glycosylation of Proteins	Bohne-Lang et al., 2005
Modeller	Comparative Protein Structure Modeling	Eswar et al., 2007
Pfam server	Protein Families Database	Finn et al., 2016
Phyre <sup>2</sup>	Protein Homology/Analogy Recognition Engine	Kelly et al., 2015
ProtParam	Calculation of physicochemical protein properties	Gasteiger et al., 2005
PSIPRED	Protein Sequence Analysis	Buchan et al., 2013
UniProt	Protein DataBase	The UniProt Consortium, 2017

### 2.2.2 Vector design

All protein constructs were expressed using the pXLG vector, (suitable for mammalian expression in human embryonic kidney (HEK) cells), which was kindly provided by Dr. David Hacker and Prof. Florian Wurm (Protein Expression Core Facility, EPFL Lausanne, Switzerland) (Figure M-1). Prior to use, the pXLG vector was modified by inserting a KpnI restriction site, immediately after the native secretion signal peptide (amino acid residues 1-17) for every DSCAM construct. In addition, a SacI restriction site was also introduced, followed by a polyhistidine tag (His<sub>6</sub>-tag) and a double stop codon at the C-terminus.

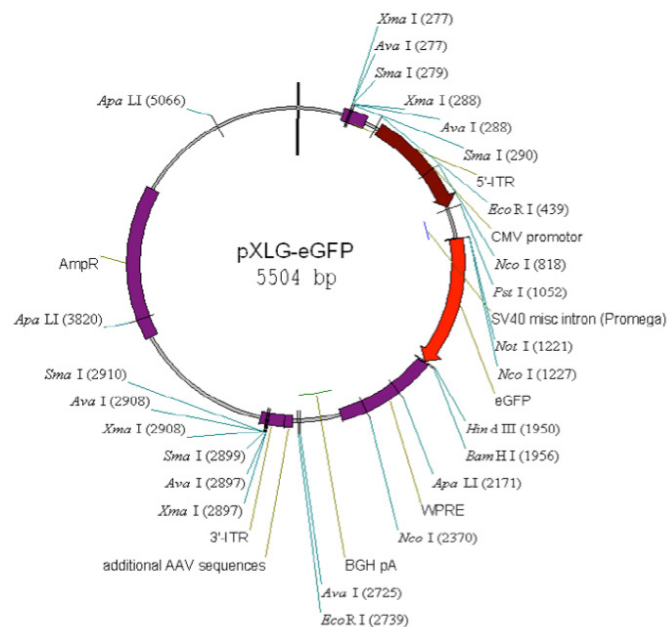


Figure M-1. Schematic representation of the pXLG vector.

## 2.2.3 Molecular cloning

### 2.2.3.1 Preparation of insert gene

The synthetic gene encoding DSCAM was received in the pUC57 vector and was ligated into the pXLG vector, using the NotI and HindIII restriction enzymes (Figure M-2). More specifically, the pUC57 vector containing the DSCAM insert gene, was amplified by means of polymerase chain reaction (PCR) using the high-fidelity phusion polymerase (Table S2) and the primers shown in Table M-3. Each primer was designed to contain the appropriate combination of restriction sites, so as to enable the generation of suitable overhangs.

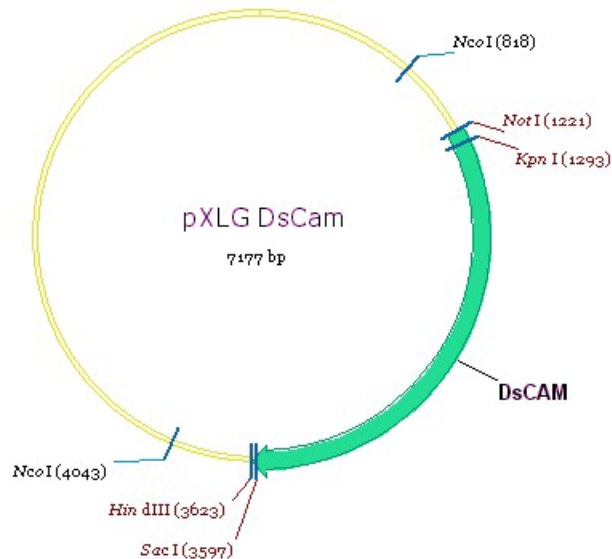


Figure M-2. Schematic representation of the pXLG vector with DSCAM.

Table M-3. Primers for preparation of the insert gene.

**Primers for DSCAM Ig1-Ig8 (native secretion signal - His<sub>6</sub>-tag)**

Forward with NotI:  
5'-AATACCGGTCATAGTTCCTGTATTTTGTGAAC-3'  
Reverse with SacI:  
5'-AGGGGTACCGAGCTCAATCTTCACAGTCA-3'

Plasmid amplification was performed using the PCR protocols described at Tables M-4 & M-5 with the necessary positive and negative controls. The polymerase was added at the last step, prior to the reaction. The total reaction volume was 50  $\mu$ l.

**Table M-4.** PCR reaction protocol for plasmid amplification.

Reagent	Volume ( $\mu$ l)
Polymerase HF reaction buffer (5x)	10
dNTPs (2 mM)	5
DNA template (50-150 ng/ $\mu$ l)	5
Forward primer (10 pmol/ $\mu$ l)	1.5
Reverse primer (10 pmol/ $\mu$ l)	1.5
Phusion HF DNA polymerase (2 U/ $\mu$ l)	0.7
Milli-Q H <sub>2</sub> O (nuclease-free)	26.3

**Table M-5.** PCR thermocycling conditions for plasmid amplification.

Step	Temperature ( $^{\circ}$ C)	Time	Repeats
Initial denaturation	98	2 min	1
Denaturation	98	30 sec	18
Annealing	63	30 sec	18
Initial extension	72	30 sec/kb	18
Final extension	72	10 min	1
Storage	4	-	-

#### 2.2.3.2 DNA digestion and ligation in restriction endonuclease cloning

The amplified product obtained by PCR, was mixed with DNA loading dye (Table S2) and was visualized under UV light on 1 % agarose gel. The remaining product was purified using the Qiagen quick PCR purification kit (Table S2) and the concentration was verified with the absorbance measurement at 260 nm with Nanodrop. The purified product and the pXLG vector were digested with the appropriate restriction endonucleases, according to the manufacturer's protocol. The digested products were loaded on an electrophoresis agarose gel and the bands, which corresponded to the pXLG vector and to the amplified insert gene, were removed carefully and were further purified using the Gel Extraction procedure from Qiagen, in accordance to the manufacturer's guidelines. Purification of the DNA product was followed by the ligation of the insert gene into the linearized vector. This process was performed using the T4 DNA ligase with a molar ratio of insert to vector DNA of 3:1 (Tables M-6 & M-7).

**Table M-6.** Restriction enzyme digestion reaction for inserts and vectors.

Reagent	Volume ( $\mu$ l)
Not1-HF (Cf= 0.4 U/ $\mu$ l)	1
Sac1-HF (Cf= 0.4 U/ $\mu$ l)	1
10x NEBuffer 2.1 (C = 1x)	5
Insert or vector DNA (Cf= 1.0 $\mu$ g)	x *
MilliQ H <sub>2</sub> O	Vt= 50

\* The amount of DNA, is related to its concentration C.  
Cf, Vt: Final Concentration and Total Volume, respectively.

**Table M-7.** Restriction enzyme ligation reaction for inserts with vectors.

Reagent	Quantity
10x T4 ligase buffer	2 $\mu$ l
Vector DNA *	50 ng
Insert DNA **	z ng ***
T4 DNA ligase	1 $\mu$ l
Milli-Q H <sub>2</sub> O	Vt= 20 $\mu$ l

\* The size of pXLG vector is 4783 bp.

\*\* DSCAM Ig1-Ig8 insert size is 2370 bp, DSCAM Ig9 insert size is 414 bp.

\*\*\* The required mass insert z (ng) = desired insert/vector molar ratio x mass of vector (ng) x ratio of insert to vector lengths.

### 2.2.3.3 Overlap extension cloning

Overlap extension cloning (OEC) is a method of ligation-independent cloning and relies on the fact that there is no requirement for restriction enzymes and *in vitro* ligation. It is based on the generation of complementary overhangs between the insert and the vector through PCR. This is achieved by designing suitable primers, which contain the specific extensions. The OEC requires two rounds of amplification. Initially, the insert gene is amplified with PCR using a primer with 5' ends complementary to the target site (Table M-5). This results in producing a megaprimer, the 3' ends of which can anneal and amplify the vector successively by overlap extension PCR. This method was used to produce the large DSCAM Ig1-Ig9 construct by combining DSCAM Ig1-Ig8 and DSCAM Ig9 constructs. DSCAM Ig1-Ig4 construct, which involved reduction of the main domains, was obtained as previously described in §2.2.3.1 and §2.2.3.2. All the primers were designed having complementary overhangs between the insert and the vector (Table S5).

#### 2.2.3.4 Agarose gel electrophoresis

Agarose gel electrophoresis is a standard technique for separating DNA fragments based on their base length (Raymond and Weintraub, 1959). Negatively charged molecules are placed into an electric field and they migrate towards the positively charged pole. The most commonly used matrix, which was also used in this study, is agarose. This method was applied for both qualitative analysis and for purification of DNA fragments (inserts or vectors).

The gel was prepared by dissolving 2 g of agarose (Table S2) in 200 ml of Tris-acetate-EDTA buffer (TAE) (40 mM Tris pH 8.5, 20 mM Acetic acid, 1 mM EDTA) to achieve 1 % solution. The mixture was homogenized in a microwave until it was dissolved. To visualize the DNA fragments under UV light, ethidium bromide was added to the solution to a final concentration of 0.2  $\mu\text{g/ml}$  before its polymerization. DNA samples were mixed with DNA loading dye (Table S2) and were then loaded to the gel along with 1kb DNA ladder, since the DNA fragment size exceeded 500 base pairs. TAE was used as a running buffer and voltage was applied at 80 V until the dye bands had migrated to approximately 75-80 % of the bottom of the gel.

#### 2.2.3.5 Transformation of *E. coli* competent cells

After ligation, the mixture was incorporated to chemically competent DH5a *E. coli* cells, using the heat-shock transformation method. The same method was used to transform every DNA construct.

The ligation mixture 1-10  $\text{ng}/\mu\text{l}$  (1 to 5  $\mu\text{l}$ ) was added to frozen ( $-80\text{ }^{\circ}\text{C}$ ) chemically competent cells (100  $\mu\text{l}$  for ligation or 50  $\mu\text{l}$  for DNA constructs) and the cells were incubated on ice for 30 min. Cell incubation was followed by a heat-shock, which was performed in  $42\text{ }^{\circ}\text{C}$  for 45 sec. The cells were further incubated on ice for 2 min for cell damage reduction. Finally, 80  $\mu\text{l}$  of SOC autoclaved medium (0.5 % w/v yeast extract, 2 % w/v bacto-tryptone, 2.5 mM KCl, 10 mM NaCl, 10 mM  $\text{MgCl}_2$ , 10 mM  $\text{MgSO}_4$ ) were added to the cells and the mixture was incubated at  $37\text{ }^{\circ}\text{C}$  for 40 min in a shaking incubator. After incubation, 100  $\mu\text{l}$  of culture were spread on a LB-agar plate, containing a suitable antibiotic (100 mg ampicillin stock, in 1:1000 dilution) and colonies appeared within 12-16 hours of incubation at  $37\text{ }^{\circ}\text{C}$ .

#### 2.2.3.6 *Plasmid preparation*

Plasmid preparation is a method used to isolate and purify plasmid DNA from bacteria. Single colonies were selected from the LB-agar transformation plates and used to inoculate 5 ml of LB broth (5 g/L yeast extract, 10 g/L bacto-tryptone, 5 g/L NaCl) with the appropriate antibiotic (ampicillin). The cultures were incubated for 12-16 hours at 37 °C in a shaking incubator and cells were harvested by centrifugation at 4000 rpm for 15 min at 4 °C. The supernatant was discarded and the pellet was used for plasmid purification with the QIAprep Spin Miniprep kit, according to the manufacturer's protocol. Plasmids were eluted in Milli-Q water and DNA quantification and purity were verified by measuring the absorbance at 260 nm and 280 nm.

#### 2.2.3.7 *Confirmation of construct*

The correct sequences of every construct were verified by DNA sequencing, performed by Eurofins MWG Operon. The resulting sequences were validated in comparison to the original theoretical sequences, using the multiple alignment-sequencing tool ClustalW2, EMBL-EBI.

#### 2.2.3.8 *Culture preparation for plasmid amplification*

Protein expression in the mammalian HEK cell line requires large amounts of plasmid DNA for transfection (2.0 mg/L). Therefore, *E. coli* competent cells (DH5a) were utilized for the plasmid amplification. The DNA plasmid of interest was used to transform DH5a cells, as previously described (§2.2.3.5) and positive single colonies were inoculated in 4L of LB media with the suitable antibiotic (ampicillin). Cultures were incubated at 37 °C in a shaking incubator. After 12-16 hours of incubation, the cultures were centrifuged and the pellet was retained for plasmid purification using the QIAplasmid Plus Gigaprep kit (Table S2), following the manufacturer's guidelines. DNA quantification and purity were verified by measuring the absorbance at 260 nm and 280 nm. The absorbance ratio ( $A_{260/280}$ ) was used to assess the purity of the DNA (for pure DNA, the ratio should be ~1.8).

## **2.2.4 Protein expression in Human Embryonic Kidney cells**

### *2.2.4.1 Preparation of HEK cells*

HEK 293T and HEK 293S cell lines were selected for transient expression in adherent cultures. Mammalian proteins were expressed and secreted in their natural environment for efficient post-translational modifications, necessary for their proper folding (Aricescu et al., 2006a; Aricescu and Owens, 2013). HEK 293T cells synthesize complex *N*-glycans, which due to their heterogeneity, might interfere with the formation of high quality diffracting crystals (Davis et al., 1993; Chang et al., 2007). Therefore, HEK 293S cell line, was used instead. These cells are characterized by the absence of *N*-acetylglucosaminyltransferase-I (GnT1) (Reeves et al., 2002), the main enzyme responsible for transferring  $\beta$ -1,2-*N*-acetyl-D-glucosamine (GlcNAc) to  $\alpha$ -1,3-asparagine linked mannose for the synthesis of complex *N*-glycans. Therefore, in its absence, these types of glycans are not synthesized, reducing the heterogeneity of the carbohydrate chains. HEK 293S line produces instead, lower molecular weight glycans such as Man<sub>5</sub>GlcNAc<sub>2</sub>. The proteins do not lose their ability of proper folding, however, they become sensitive to enzymatic treatments with endoglycosidases, such as endoglycosidase H (Davis et al., 1993; Butters et al., 1999; Chang et al., 2007).

### *2.2.4.2 Cell maintenance*

Cells were grown in polystyrene culture flasks with a surface area of 175 cm<sup>2</sup> in an incubator (37 °C, 5 % CO<sub>2</sub> atmosphere) (Table S1) and were maintained in 30 ml Dulbecco's modified eagle's medium (DMEM), containing 2 mM L-Glutamine, 1x non-essential amino acids and 10 % fetal calf serum (FCS) (Table S2). During cell growth, cells were subcultured before reaching their full confluence (80-90 %) in an effort to maintain their proliferative phenotype. Cell passaging was performed by aspirating culture medium (DMEM) on the cells and residual FCS was removed after a washing step with 10 ml of PBS. Cells were detached from the surface of the flask with 3 ml 1x Trypsin-EDTA (Table S2) and were subsequently incubated for ~3 min at room temperature (RT). The trypsinization reaction was quenched with 7 ml of culture media (DMEM) containing 10 % of FCS and each culture flask was passaged into new flasks (in a dilution 1:5) containing 25 ml of DMEM with 10 % FCS. An assay was then performed to determine cells viability. Cells were counted before transfection, using a hemocytometer and the quality was assessed by visualization



using 0.4 % trypan blue stain dye (Table S2), since viable cells membrane is impermeable to this type of dyes. Cultures with less than 90 % viability were discarded.

#### 2.2.4.3 Protein transfection and large-scale expression

Large-scale cultures were performed in expanded surface polystyrene roller bottles with a surface area of 2125 cm<sup>2</sup> (Table S1). ¼ of cells (with 90 % confluence), originally harvested from T175 flasks (as it was described in §2.2.4.2), were re-suspended in 7 ml DMEM with 10 % FCS and were added to 250 ml total culture medium per roller bottle. Large-scale cultures were incubated prior to transfection for 72 h, under a gentle rotation of 0.8 rpm/min. Transient gene expression was performed by chemical transfection using the branched cationic polymer polyethylenimine (PEI) (Table S2). The latter forms complexes with negatively-charged DNA molecules that bind to anionic cell surface residues, resulting in endocytosis. The amine protonation (by PEI) causes a decrease of the osmotic potential and leads to the release of the polymer-DNA complex in the cytoplasm, allowing the DNA molecules to dissociate and diffuse into the nucleus (Nimesh and Chandra, 2008). The transfection protocol involved PEI dilution to a final concentration of 1 mg/ml (pH 7.4) with a DNA to PEI molar ratio of 1:2. For large-scale expression, PEI was incubated with the plasmid of choice (500 ng DNA for 250 ml culture) in 50 ml DMEM (without addition of FCS). Incubation for 10 min was followed by the replacement of culture media with 200 ml fresh DMEM that contained 2 % FCS. Cultures were further incubated at 37 °C in low speed (0.8 rpm/min) in 5 % CO<sub>2</sub> atmosphere and the protein of interest was secreted to the medium. After 4 days, the cells were discarded using standard filter papers (Table S1). Secreted proteins in the culture media were preserved for long storage periods at 4 °C, using 1mM NaN<sub>3</sub> for microbial growth prevention. Protein expression was verified using Western blot analysis (§2.2.6.2).

### 2.2.5 Protein purification

Protein purification was conducted at 4 °C and all protein constructs were handled with extra precaution on ice throughout the experiments. Classical liquid chromatography techniques were employed taking advantage of the sample properties. These involved:

(i) Immobilized metal affinity chromatography (IMAC), firstly introduced by Porath and coworkers (Porath et al., 1975), that was used to purify polyhistidine-tagged proteins or proteins containing cysteine or tryptophan residues, utilizing the natural binding tendency of certain amino acid side-chains to metal ions (e.g. Ni<sup>+2</sup>, Co<sup>+2</sup>, Zn<sup>+2</sup>, Cu<sup>+2</sup>). The desired tagged protein is bound on a resin bearing immobilized metal ions and then is eluted by either pH or imidazole gradient. The choice of immobilized metal ion is related to its affinity and specificity for the binding proteins.

For DSCAM protein constructs (His<sub>6</sub>-tag located at C-terminus), Nickel sepharose Excel beads (Table S2) were used. The advantage of that resin relies on the fact that Ni<sup>+2</sup> is strongly bound to a novel chelating ligand with high capacity and minimum ion leakage. Thus, no additional dialysis of protein media is necessary prior to the purification step. One Liter of medium, for all DSCAM constructs, was incubated at 4 °C over night with 4 ml Ni<sup>+2</sup> Excel beads, which were equilibrated with 60 ml of Washing Buffer I prior to use. The protein-infused beads were then transferred to a Bio-Rad glass gravity column (Table S1) for 30-minute incubation at 4 °C with 50 ml of Washing Buffer I. After a two-step washing (2 column volumes (CV) of Washing Buffer I and 1 CV of Washing Buffer II), the protein was eluted into 6x 2 ml fractions (Elution Buffer) until no further protein was detected (UV measurement at 280 nm). The molar extinction coefficient for DSCAM Ig1-Ig9 is  $\epsilon=103140$  (L.mol<sup>-1</sup>.cm<sup>-1</sup>) or 1.07 (for 0.1 % absorbance in 1 g/L), for DSCAM Ig1-Ig8 is  $\epsilon=93170$  (L.mol<sup>-1</sup>.cm<sup>-1</sup>) or 1.05 (for 0.1 % absorbance in 1 g/L) and for DSCAM Ig1-Ig4 is  $\epsilon=48820$  (L.mol<sup>-1</sup>.cm<sup>-1</sup>) or 1.12 (for 0.1 % absorbance in 1 g/L). Buffers used for IMAC purification are summarized in Table M-8.

Table M-8. Composition of main buffers used for IMAC purification of DSCAM.

Washing Buffer I	Washing Buffer II	Elution Buffer
20 mM Phosphate pH 7.4 500 mM NaCl 1 mM TCEP	20 mM Phosphate pH 7.4 500 mM NaCl 1 mM TCEP 50 mM Imidazole	20 mM Phosphate pH 7.4 500 mM NaCl 1 mM TCEP 500 mM Imidazole

(ii) Size exclusion chromatography (SEC) that utilizes differences of proteins regarding their hydrodynamic volume. The protein solution migrates through a column packed with a matrix made of porous material of spherical particles and separation is achieved due to different elution time of each protein, based on their size differences.

The elution fractions, resulting from IMAC purification, were filtered with a 0.22 $\mu$ m membrane prior to SEC purification and were loaded on a Hiload Superdex 200 prep grade column using the ÄKTA liquid chromatography purification system. The column was equilibrated with a different SEC buffer corresponding to different protein constructs (Table M-9). Purity and identification of the protein were determined by SDS-PAGE and Western blot analysis and the final concentration via absorbance measurements at 280 nm.

**Table M-9.** Composition of main buffers used for SEC purification of DSCAM.

<b>Protein</b>	<b>SEC Buffer</b>
DSCAM Ig1-Ig8	50 mM Citrate pH 6.0, 150 mM NaCl, 2 mM DTT
DSCAM Ig1-Ig4 & DSCAM Ig9	50 mM MES pH 6.0, 250 mM NaCl, 1 mM DTT
DSCAM Ig1-Ig9	50 mM MES pH 6.0, 150 mM NaCl, 1 mM DTT

(iii) Ion exchange chromatography (IEC) that relies on differences of proteins in their net surface charge in relation to pH, a property unique for every protein. Protein-binding to resin material of the column, is driven by the ionic interaction between the oppositely charged protein material (mobile phase in low ionic strength solution) and the material of the column. The sample is eluted by disrupting this interaction, using a solution of high ionic strength in an increasing gradient.

IEC was used for DSCAM Ig1-Ig9 in order to prepare the protein for negative staining electron microscopy studies. More specifically, anion IEC was used after crosslinking of DSCAM with glutaraldehyde solution in order to separate the crosslinked from the non-crosslinked population. A MonoQ 5/50 GL column was equilibrated with 10 CV of Buffer A, 10 CV of Buffer B and 10 CV of Buffer A (or until the baseline, conductivity and eluent pH were stable) (Table M-10). The sample was injected to the column and was eluted with a gradient volume of 60 CV and an increasing ionic strength up to 1 M NaCl. The column was washed with 5 CV of Buffer B to elute any remaining bound material and it was re-equilibrated with 10 CV

of Buffer A. The eluates were analyzed with SDS-PAGE to determine whether protein separation was achieved.

**Table M-10.** Composition of buffers used for IEC purification of DSCAM Ig1-Ig9.

<b>Protein</b>	<b>Buffer A</b>	<b>Buffer B</b>
DSCAM Ig1-Ig9	50 mM MES pH 6.0 10 mM NaCl, 1mM DTT	50 mM MES pH 6.0 1 M NaCl, 1mM DTT

## 2.2.6 Protein detection and purity assessment

### 2.2.6.1 SDS-PAGE

Quality and relative molecular mass of every protein sample, was assessed with sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), a widely-used technique (Laemmli, 1970) for protein separation and analysis. The polypeptide chains bind to SDS, which is an anionic detergent that dissociates the tertiary structure of the protein coating it with a uniform negative charge, which is approximately proportional to the protein's molecular mass. The protein samples are heated to 95 °C in the presence of reducing agents and proteins are transformed to linear molecules, which can migrate in an electric field according to their charge-to-molar ratio.

The Mini PROTEAN Tetra Cell (Table S1) system was used for electrophoresis for all protein constructs. Aliquots from every protein sample (9  $\mu$ l) at each purification stage were mixed with 3  $\mu$ l of 4x NuPAGE LDS sample buffer supplied with 10 %  $\beta$ -mercaptoethanol. Protein samples were denatured at 95 °C for 5 min, centrifuged for 60 sec and loaded on a Bis-Tris gel. Gel composition is shown in Table M-11. The amount of acrylamide for every gel was adjusted according to the molecular weight (MW) of the protein.

Table M-11. Composition of Bis-Tris gels.

2 gels 10ml	15 %	12 %	10 %	8 %	Stacking Gel
5x gel buffer*	2 ml	2 ml	2 ml	2 ml	0.8 ml
40 % acrylamide	3.8 ml	3 ml	2.5 ml	2 ml	0.5 ml
MilliQ	4.2 ml	4.9 ml	5.4 ml	5.9 ml	2.7 ml
20 % APS	25 $\mu$ l	25 $\mu$ l	25 $\mu$ l	25 $\mu$ l	25 $\mu$ l
TEMED	12.5 $\mu$ l	12.5 $\mu$ l	12.5 $\mu$ l	12.5 $\mu$ l	12.5 $\mu$ l

\* 200 ml of 5x gel buffer: 74.7 g of Bis-Tris methane pH 6.5-6.8 with HCl.

Gels were run at constant voltage (180 V) using the running buffers MES or MOPS, for low and high MW proteins, respectively (Table M-12).

Table M-12. Composition of running buffers used for SDS-PAGE.

Components	20x MES (500 ml) 5-20 kDa proteins	20x MOPS (500 ml) proteins > 20 kDa
	97.6 g MES	104.6 g MOPS
Tris base	60.6 g	60.6 g
SDS	10 g	10 g
EDTA	3 g	3 g

Gels were stained using the InstantBlue coomassie based stain (Table S2). The protein bands appeared within 15 min, with no additional destaining and washing steps.

#### 2.2.6.2 Western blot

Protein immunoblot is a widely used technique for protein quality assessment based on the use of antibodies specific to target proteins (Towbin et al., 1979). After protein separation using SDS-PAGE, the gel was equilibrated for 5 min in Transfer Buffer (Table M-13) to remove the excess salt from electrophoresis and it was transferred to a nitrocellulose membrane using the Mini Trans-Blot Cell (Table S1). The Transfer Running Buffer was pre-chilled at 4 °C and the blotting cell was kept on ice during the electrophoretic transfer (constant voltage 100 V for 60 min) to decrease the heat produced. The membrane was subsequently washed with 3x 20ml PBS-T Buffer and then it was blocked for 60 min at RT with 5 % non-fat milk (in PBS-T Buffer) to reduce the non-specific binding sites of the primary antibody. After an additional washing step, the membrane was incubated at RT with a penta-His primary antibody (1:2000 dilution in PBS-T with 0.05 % non-fat milk) for 60 min, followed by the incubation of the secondary antibody (1:5000 dilution in PBS-T with 0.05 % non-fat milk) at RT for 30 min (Table S3). Several washing steps were performed, between the antibody incubation periods, for preventing non-specific binding. Detection of the proteins was performed with the SuperSignal West Pico Chemiluminescent Substrate (Table S2). The solution was prepared by mixing equal parts of stable peroxide substrate reagent with luminol solution (Table S2). The membrane was covered with 5 ml of that mixture and it was exposed using the GelDoc system, with an initial exposure time of 30 sec up to 5 min, depending on the protein concentration.

Table M-13. Composition of buffers used for Western blot.

10x Transfer Buffer	10x PBS pH 7.4	PBS-T
25 mM Tris-HCl pH 8.3	100 mM Na <sub>2</sub> HPO <sub>4</sub> ·7H <sub>2</sub> O	1x PBS + 0.05 % w/v Tween 20
192 mM glycine	18 mM KH <sub>2</sub> PO <sub>4</sub>	
20 % v/v methanol	1.37 M NaCl	
	27 mM KCl	

## 2.2.7 *Biophysical methods*

### 2.2.7.1 *MALDI-TOF analysis*

Mass spectrometry (MS) is an analytical technique for detection and molecular weight (MW) determination of biomolecules (Hoffmann and Stroobant, 2007). Matrix-assisted laser desorption/ionization (MALDI) is a soft ionization technique that leaves the protein molecules intact. The protein, either in solution or in crystalline form, is mixed with a compound (matrix solution) under denaturing conditions and the mixture is applied on a stainless steel plate that is exposed at RT until dehydration is achieved. The dried material is placed in the mass analyzer chamber where the sample absorbs UV light (337 nm) in a short pulse duration, leading to its partial evaporation and protonation from the matrix compounds. The protonated molecules are accelerated in an applied electric field and are further analyzed, according to their time of flight (TOF) by estimating their mass-to-charge ratio ( $m/z$ ).

MALDI-TOF was the method of choice for protein assessment and MW determination in this study and was performed by the Sample Preparation and Characterization Facility (SPC, EMBL-Hamburg). 2  $\mu$ l of protein sample in solution at 1 - 10  $\mu$ M concentration were mixed with the matrix solution (50 % acetonitrile, 0.1% v/v trifluoroacetic acid) in a 1:1 or 1:2 ratio. The mixture was applied on a sample plate until it was crystallized. For samples where the MW of a crystallized protein had to be determined, a single crystal was removed from the crystallization drop and was introduced to a fresh-5  $\mu$ l-drop containing only the crystallization conditions. The crystal was subsequently washed with 3x 5  $\mu$ l of reservoir solution, to remove any remaining protein solution around the crystal. The crystal was then washed twice with 5  $\mu$ l of acetonitrile, in order to reassure that non-crystallized protein and PEG will not interfere with the MW determination of the protein crystal (Nettleship et al., 2005). In the final step, the liquid in excess was removed by wicking (Table S1) and the crystal was dissolved in 15  $\mu$ l buffer of 20 mM Tris, 150 mM NaCl, 8 M urea and it was transferred to the sample plate until dehydration was achieved. All protein analyses were performed by the Voyager DE-STR MALDI-TOF mass spectrometer (Table S1).

### 2.2.7.2 *Thermal shift assay (Thermofluor)*

Thermofluor is a thermal denaturation assay, used for preliminary characterization of the protein to identify its stability and solubility under different conditions (buffers, pH, salts, ionic strength, ligands) by employing fluorescence

detection (Pantoliano et al., 2001). Thermofluor can also be used for optimization of existing crystallization conditions and detection of additives that might increase protein stability (Phillips and de la Peña, 2011). A fluorescent dye binds to hydrophobic residues and the emitted energy, as the protein unfolds, is monitored by a qPCR machine (Real-Time PCR). The melting temperature of the protein ( $\Delta T_m$ ) is then determined by the unfolding curves. The conditions that stabilize the protein will result in curves with an increased melting temperature.

In the frame of this project, protein samples were added in a PCR plate with 96 different conditions, designed by the SPC staff members at EMBL-Hamburg (Boivin et al., 2013). Each well contained also the components shown in [Table M-14](#) and the thermal denaturation analysis was performed in iCycler iQ system with an optical module for fluorescence detection ([Table S1](#)). The plate was successively heated from 5 °C - 95 °C with a temperature increase of 1 °C/min. Data were analyzed using GraphPad Prism 6.0 (GraphPad Software, Inc).

[Table M-14](#). Thermofluor 96-well assay.

<b>Components</b>	<b>Volume (<math>\mu</math>l)</b>
ddH <sub>2</sub> O	16
5x buffer	5
62.5x SYPRO orange	2
Protein sample (20 - 100 $\mu$ M)	2

### 2.2.7.3 FPLC-RALS/RI/UV

Protein quality assessment of oligomeric state and MW determination were performed by fast performance liquid chromatography (FPLC) coupled with static light scattering (SLS). The technique combines right-angle light scattering (RALS) measurements along with refractive index (RI) and UV absorbance (280 nm), using the Malvern Instrument ([Table S1](#)). Protein samples for all constructs (~8 mg/ml, 100  $\mu$ l) were filtered through a 0.22  $\mu$ m membrane filter in order to remove potential aggregates and were loaded on an analytical column (Superdex 200 10/300 GL) for SEC. Fractions were collected with a flow rate of 0.4 ml/min in RT. The buffers used for every construct were the same as those for SEC ([Table M-9](#)). The intensity of scattered light by protein sample was measured at RT and MW determination followed. MW of the eluted sample was calculated by using OmniSEC software (Malvern's OmniSEC GPC/SEC multidetector system) according to either



concentration measurements (derived from base-line RI corrections) or UV measurements (derived from base-line corrected RALS intensities), using the following formulas (Follmer et al., 2004; Ma et al., 2015):

$$\text{RALS} = c(\text{dn}/\text{dc})^2 \times \text{MW} \times K_{\text{RALS}}$$

$$\text{RI} = c(\text{dn}/\text{dc}) \times K_{\text{RI}}$$

$$\text{UV} = c\varepsilon K_{\text{UV}}$$

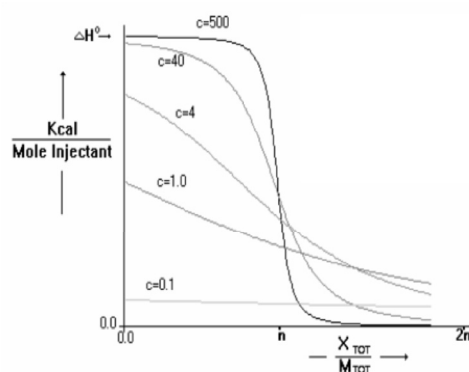
\*  $K_{\text{RI}}$ ,  $K_{\text{RALS}}$ ,  $K_{\text{UV}}$ : instrument calibration constants determined using a protein standard (i.e. Bovine serum albumin (BSA))

\*  $\text{dn}/\text{dc} = 0.185 \text{ ml/g}$ : refractive index increment for unmodified protein

\*  $\varepsilon$ : extinction coefficient ( $A = 0.1 \% \text{ mg/ml}$ )

#### 2.2.7.4 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is a technique for studying the interactions of molecules upon binding to their counterparts, by measuring the transferred heat either in endothermic or in exothermic processes. The reaction between two components takes place in an adiabatic system in constant temperature. The molecule of interest is kept in a stirred cell and the ligand or the binding partner is titrated through a syringe. Upon binding, the energy released or absorbed from the reaction of the two components, is measured by a microcalorimeter that detects the temperature change. The binding affinity is then quantified as a function of the total energy of the reaction, which is in direct proportion to the amount of complex formation. Upon binding, the protein in the cell is saturated with its counterpart and the heat change decreases until reaching the background dilution heat. The parameters that can be determined, after data analysis, are the stoichiometry of the reaction ( $n$ ), the binding constant ( $K_d$ ), enthalpy ( $\Delta H$ ) and entropy ( $\Delta S$ ) (Freyer and Lewis, 2008; Duff et al., 2011).



The experimental data fit to a mathematical model in order to maximize the confidence of the calculating parameters. This is referred as a *c* window ( $10 < c < 500$ , where *c* is a dimensionless constant) and when  $20 < c < 100$ , the values obtained are very accurate.

$$c = nM[\text{tot}]/K_d$$

- \* n: stoichiometry of reaction
- \* M[*tot*]: molar concentration in the cell
- \* *K<sub>d</sub>*: dissociation constant

ITC experiments were conducted for studying DSCAM Ig1-Ig9:netrin<sub>VIV</sub> interactions. The assays were performed in a VP-system at 25 °C in an identical buffer for both proteins (Buffer 1 and Buffer 2 were used in two independent experiments). DSCAM was titrated to a netrin solution using 30 injections of 3 μl with time intervals of 360 sec between the injections. Data were processed with Origin7 software (MicroCal) (Table M-15).

Table M-15. ITC experimental set-up and buffer composition.

Components	Concentration	Volume (ml)
Netrin <sub>VIV</sub> (Cell)	10 μM	1.4
DSCAM Ig1-Ig9 (Syringe-titrated partner)	100 μM	0.25
Buffer 1	50 mM MES pH 6.0 150 mM NaCl 1mM (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> 1 mM DTT, 2 mM CaCl <sub>2</sub>	
Buffer 2	1x PBS pH 8.0	

## 2.2.8 Structural characterization

### 2.2.8.1 Small-angle X-ray scattering

Small-angle X-Ray scattering (SAXS) is a method for biomolecular shape and size determination (Svergun and Koch, 2003) that can provide structural information for proteins and their conformations (at low-resolution 50 Å - 10 Å) complementary to the one obtained from other methods used for structure determination. Structural characterization of proteins using SAXS is performed in solution where the molecules are studied in their native form close to their physiological environment, overcoming the crystallization stage known to be a bottleneck in X-ray crystallography. Comparison of SAXS with NMR, which also involves structural studies of macromolecules in solution, shows that SAXS, although it is of lower resolution, is widely used for large proteins and multicomponent macromolecular complexes without any limitations imposed by protein size. SAXS is also combined with X-ray crystallography, Electron Microscopy, NMR and computational methods resulting in hybrid modeling approaches.

#### a. SAXS data collection and analysis

SAXS data were collected for DSCAM Ig1-Ig4, DSCAM Ig1-Ig8, DSCAM Ig1-Ig9 and for netrin<sub>IV</sub> in association with DSCAM Ig1-Ig9. Sample purity and monodispersity are essential for SAXS data analysis and structural modeling. Therefore, after SEC purification (Table M-9), the eluted samples were evaluated by SDS-PAGE and prior to use, were filtered with 0.22 μm membrane filters to discard possible aggregates. The protein concentration was determined at 280 nm ( $\epsilon$  in §2.2.5i) by Nanodrop. For each sample a range of concentrations were tested (Tables S6-S9). Protein samples (30 μl) were loaded into the automatic sample chamber (10 °C) and were exposed to a highly monochromatic focused X-ray beam ( $\lambda = 1.24$  Å), while flowing through a quartz capillary (P12, PETRA III, EMBL-Hamburg, DESY, Germany). Scattered X-rays were recorded on a 2D Pilatus detector (Pilatus 2M, Dectris). The sample-to-detector distance was 3.1 m, covering a momentum transfer range of  $0.008 \text{ \AA}^{-1} \leq s \leq 0.457 \text{ \AA}^{-1}$  ( $s = 4\pi \sin(\theta)/\lambda$ ,  $2\theta$ : scattering angle and  $\lambda$ : X-ray wavelength). Each measurement was taken as an average of 20 successive frames with 50 ms exposure time in order to assess potential radiation damage. Comparison of the frames showed no significant change, thus no radiation damage was observed for these samples.

Detector images were radially averaged to yield 1D scattering profiles and were normalized to the intensity of the transmitted beam and to the exposure time. Sample contributions were averaged and corrected in relation to the solvent (buffer subtraction) and data collected were extrapolated to infinite dilution. The overall parameters were evaluated using the graphical package PRIMUS (Konarev et al., 2003). More specifically, the radius of gyration ( $R_g$ ) and the forward scattering intensity  $I(0)$  were determined from the scattering profile using Guinier approximation (Guinier, 1939) where for very small scattering angles ( $s \leq 1.3/R_g$ ), the intensity is represented by the following formula:

$$I(s) \cong I(0) \cdot e^{-\frac{1}{3}(sR_g)^2}$$

The maximum particle dimension,  $D_{\max}$  was obtained from the interatomic distance distribution function,  $p(r)$  and was calculated with the software GNOM (Svergun, 1992). The molecular mass (MM) of every protein construct was estimated by comparison of the extrapolated forward scattering to that of a reference protein (BSA). A second independent molecular mass estimation was calculated using the Porod equation (Porod, 1982) which estimates the excluded volume of the hydrated particle,  $V_p$  ( $s_{\max}=7.5/R_g$  to remove any undesired scattering contribution from the internal structure). The molecular mass of the particle can be derived from  $V_p$ -value in  $\text{\AA}^3$  and it is approximately 1.7 times the MM of the protein (in Dalton) (Petoukhov et al., 2012). Superposition of low-resolution models with X-ray crystal structures was performed with SUPCOMB (Kozin and Svergun, 2001).

#### *b. Ab initio shape determination and model generation*

Low-resolution shape bead models were determined for all constructs with the *ab initio* modeling method, using software DAMMIN (Franke and Svergun, 2009). DAMMIN uses simulated annealing to represent proteins as bead models where the theoretical scattering of the proteins in solution fit optimally the experimental data. Several independent modeling runs are performed and DAMAVER is used to cluster and determine the averaged reconstruction (Volkov and Svergun, 2003).

#### *c. Rigid body modeling*

Rigid body models for each data set were constructed based on high-resolution models from the corresponding crystallographic structures. Real-space arrangement of subunits was refined against the experimental SAXS data using Monte-Carlo

stimulations with software CORAL (Petoukhov et al., 2012). Models were initially reconstructed in P1 to assess the robustness of model generation, followed by P2 symmetry for DSCAM Ig1-Ig8 and DSCAM Ig1-Ig9. For each data set, ten independent model reconstructions were performed and the most representative model of the ensemble was identified using DAMAVER (Volkov and Svergun, 2003).

#### *2.2.8.2 Electron microscopy*

Electron microscopy (EM) is a widely used method that utilizes a beam of electrons as a light source, to obtain structural information of macromolecules without the need of a crystalline sample. Different approaches have been used for sample preparation and specimen analysis (e.g. cryo-fixation, negative staining, conductive coating, etc) (Frank, 2006). Cryo-EM, in particular, provides structural information at medium to high-resolution in order to produce an accurate 3D reconstruction model of the biological object. Although efforts have been made to simplify the data collection and automated processing, the method has some limitations imposed by the validation tools for evaluating the quality of the density maps (Cheng et al., 2015). Sample homogeneity is a critical factor for molecular averaging and classification (single-particle analysis) by EM. Conventional methods employed for sample preparation and evaluation, such as protein purification and SDS-PAGE, are insufficient to provide information about the sample consistency. Therefore, negative staining EM is in this case essential to assess the sample's homogeneity and to identify different protein conformations. Additionally, negative staining introduces some benefits in analyzing heterogeneous populations in comparison with cryo-EM where vitrified molecules may adopt random conformations and orientations (Frank, 2006).

##### *a. Negative staining*

Negative staining is a tool, where molecules or individual cell organelles, are embedded in a heavy metal solution, in RT, and they are visualized by exploiting the image contrast between the sample and the background. With the implementation of single-particle analysis and classification, useful information can be extracted for the shape and size of the macromolecule (Ohi et al., 2004). Samples in suitable buffers (phosphate buffers should be avoided to prevent stain precipitation on the grid) are stained with a solution and due to contrast enhancement, smaller molecules can be easily detected, in comparison with cryo-EM methods (Burgess et al., 2004; Ohi et

al., 2004). In addition, negative staining is used to assess sample purity and heterogeneity. One should take into consideration that negative staining also bears some limitations. During sample staining, the molecules are dehydrated and they might be visualized as flattened or distorted, making it difficult to deduce their physiological function. In addition, negative staining is not recommended in cases that high-resolution analysis is required for obtaining structural information. However, it is a suitable method for revealing structural features at 15 - 20 Å resolution for molecules with a MW of at least 110 kDa (Jawhari et al., 2006).

Negative staining was used to investigate DSCAM dimerization and to determine the binding mode between DSCAM Ig1-Ig9 and netrin<sub>VIV</sub>. Specimens were prepared following the conventional negative staining protocol (Ohi et al., 2004) where 2.5 µl of protein solution (~5 - 10 µg/ml) was absorbed in a glow carbon-coated copper grid. The grid was washed with deionized water and the specimens were stained with 0.75% uranyl formate (or uranyl acetate and ammonium molybdate) at RT. Raw sample images were taken with a Tecnai T12 electron microscope (operated at 120 kV) using low-dose procedures and they were recorded on a Gatan US4000 CCD camera (magnification x71,138 and defocus value of ~1.0 µm). All images were binned over 2x2 pixels to obtain a pixel size of 4.16 Å on the specimen level and they were subjected to single-particle analysis where two-dimensional reference-free alignment and classification averages were obtained by ISAC (Yang et al., 2012).

#### *b. Protein crosslinking*

Mild chemical crosslinking is used to reduce the heterogeneity of the sample and to visualize different protein conformations. Although it is a widely used approach, there is a propensity for artifacts introduced by crosslinking. These include, positional fixation of the most favorable or most compact conformation of the protein or clustering of flexible regions, possibly leading to results with non-biological relevance (Cheng et al., 2015; Thompson et al., 2016).

Mild protein crosslinking was performed in glutaraldehyde solution through gravity (PD-10 columns) (Table S1) or size exclusion columns ('on-column' protocol) (Shukla et al., 2014). For this purpose, EM-grade glutaraldehyde (Table S2) was used to crosslink the primary amines of DSCAM Ig1-Ig9 in an effort to stabilize the dimer interactions among its Ig domains. Initially, glutaraldehyde in a range of 0.1-0.8% v/v was tested for different protein concentrations and buffer solutions to identify the optimum conditions for protein crosslinking. For the PD-10 protocol, the

columns were equilibrated with Stabilization Buffer (SB) (3x 3 ml). Equilibration was followed by the addition of 400  $\mu$ l SB (until reaching the bed volume). Successively, on the column were added: 400  $\mu$ l glutaraldehyde, 400  $\mu$ l SB, 400  $\mu$ l protein, 2x 400  $\mu$ l SB. The protein was eluted with 4x 0.9 ml buffer in small fractions of 30  $\mu$ l and the crosslinked protein was detected by SDS-PAGE. Details of the protocol applied are described in [Table M-16](#).

Crosslinking was also performed in a batch-method using 1 ml of protein (1 mg) in both Stabilization Buffers with 5  $\mu$ l of 2.3 % glutaraldehyde. The solution was incubated at 37 °C for 5 min and the reaction was quenched with 100  $\mu$ l of 1 M Tris-HCl pH 8.0. The crosslinked protein was assessed by SDS-PAGE.

[Table M-16](#). Protein crosslinking of DSCAM Ig1-Ig9 in PD-10 columns.

Components	Quantity	Volume ( $\mu$ l)
Glutaraldehyde (% v/v)	0.1, 0.2, 0.4, 0.8	400
Stabilization Buffer 1	50mM MES pH 6.0, 1mM CaCl <sub>2</sub> 150 mM NaCl, 1mM DTT	400
Stabilization Buffer 2	1x PBS pH 7.4	400
DSCAM Ig1-Ig9	1 mg/ml	400

### 2.2.8.3 X-ray protein crystallography

#### a. Protein crystallization

X-ray crystallography is a widely used method for the 3D structure determination of macromolecules, which at atomic resolution gives insights into the amino acid interaction network and in most cases reveals the biological function and the role of the macromolecules in the cell. X-ray crystallography requires a periodic arrangement of the protein molecules to ensure enhancement of the diffracted beam for enabling the signal to be recorded by the detector. This can be achieved by forcing the protein molecules to be arranged in a crystalline form. This process is rather challenging, considering that protein amino acids forming the polypeptide chain, adopt different conformations that might affect packing interactions of the molecules under crystallization (Deller et al., 2016). In addition, several parameters such as the sensitivity of protein samples to various conditions (e.g. temperature, ionic strength), the fragility of protein crystals compared to inorganic ones, due to weak interactions formed between the flexible amino acids and water molecules in

the crystal and the fact that there are no 'universal' crystallization conditions for a given protein, make crystal growth a demanding process (McPherson 2004; Rhodes, 2000). The quality of the protein material, in terms of homogeneity, solubility and purity is also a critical factor in order to obtain diffracting crystals. Thus, protein purification protocols should be carefully designed and the suitability of protein samples, prior to crystallization, should be verified by biophysical characterization.

Successful crystallization involves the identification of conditions that would bring protein molecules from the soluble phase to the supersaturation zone where crystals can nucleate and grow. Transition between the phases is induced by modifying a number of parameters such as protein concentration, type and concentration of precipitating agents, ionic strength, pH, temperature, etc (Chayen and Saridakis, 2008). Throughout the years several different methods have been developed for increasing the success rate of crystallization attempts. In this context, efforts for high-throughput screening and miniaturization of protein sample volumes have been made with emphasis on both the crystallization medium (crystallization kits) and the equipment used (robotic devices).

#### *b. Crystallization methods*

A number of methods and set ups are employed in order to achieve crystal growth (e.g. vapor diffusion, dialysis, batch/microbatch, etc). The most commonly used is the vapor diffusion method with sitting or hanging drops. A drop containing the protein with the mother liquor, in an isolated system, is suspended on a cover slip over the reservoir solution (hanging drop) or is positioned on a shelf within the well (sitting drop). The precipitant concentration in the reservoir solution and in the sample drop is different (the reservoir concentration is higher) and vapors from the sample drop diffuse into the reservoir until equilibrium is achieved. Protein concentration increases in the sample drop and depending on the combination of the precipitants used, crystallization might occur (Abts et al., 2012). In the batch method, the protein solution and the precipitant agents are mixed directly and are kept in a sealed isolated environment. The concentration of precipitants along with the protein remains constant throughout the experiment (Rayment, 2002). In some cases, to facilitate crystal growth the nucleation step is induced by adding small crystalline seeds at various dilutions. Crystal seeding is a tool that is also used with other techniques such as vapor diffusion, and it allows separating nucleation and crystal growth. In this method, seeds taken from previously nucleated crystals, are



introduced into a new crystallization drop in various dilutions (Bergfors, 2003). For microbatch crystallization (a variation of the batch method), a thin layer of oil (paraffin oil or a mixture of both paraffin and silicone oils) covers the drops, which are suspended in a microwell. The oil allows slow evaporation of water, which leads to concentration increase for both protein and precipitant agents, yielding to protein crystallization (Brumshtein et al., 2008; Chayen et al., 1990). Dialysis is also used for protein crystallization as a variation of the vapor diffusion method. This involves separation of the protein from the crystallization medium by a semi-permeable membrane that allows only ions and buffer components to cross the membrane (Chayen and Saridakis, 2008).

### *c. Crystallization of DSCAM*

Sparse-matrix screening involves usage of diverse precipitant agents to explore crystallization conditions and is often used as the first step in protein crystallization. The sparse-matrix screens are based on existing conditions from previously determined crystallographic structures in order to reveal a relationship among different parameters (Carter and Carter, 1979; Jancarik and Kim, 1991). In addition to 'randomly' probing sparse matrix screens, systematic grid screens on pH versus specific precipitants (such as ammonium sulphate grid) can be also used to identify initial crystallization conditions or to optimize a hit from a sparse-matrix screen.

Protein samples of DSCAM Ig1-Ig4, DSCAM Ig1-Ig8 and DSCAM Ig1-Ig9 were concentrated to 10, 8 and 10 mg/ml (final concentration), respectively and centrifuged for 10 min at 10,000 rpm (4 °C) to remove potential aggregates. Pre-crystallization tests (PCT, Hampton Research) were performed to estimate approximately the starting concentration for crystallization experiments (Table S2). A variety of commercial screens from Qiagen, Molecular Dimensions and Hampton Research (Table S2) were used to identify the initial crystallization conditions. Experiments were conducted at the high-throughput crystallization facility (SPC-HTX, EMBL-Hamburg) (Boivin et al., 2015) using 96-well Greiner or iQ plates (Table S1). The final drop volume was 300 - 400 nl and the reservoir solution was mixed with the protein in different volume ratios (protein to reservoir solution 1:1, 1:2 and 2:1). The mixture was equilibrated against 50  $\mu$ l reservoir solution applying the sitting-drop vapor diffusion method. Plates were set using the Mosquito-LCP crystallization robot (Table S1) and were stored at 19 °C, where they were monitored periodically by automated image acquisition (Rock Imager by Formulatrix). Optimization and fine-

tuning of the crystallization conditions using rationally designed custom-made screens, was followed, based on the initial crystal hits, using the hanging-drop vapor diffusion method (in 24-well Linbro plates). A variation of different crystallization components and parameters were tested, such as temperature (4 - 20 °C), pH, drop volume and shape. For DSCAM Ig1-Ig4 and DSCAM Ig1-Ig8, optimization trials were also made with emphasis on changing the kinetics of the diffusion rate between the drop and reservoir solution. For this purpose, the vapor diffusion method under oil was used. A combination of paraffin and silicon oil was mixed in 50:50 and 40:60 (silicon to paraffin) volume ratios and the mixture (200  $\mu$ l - 600  $\mu$ l) was applied on top of the reservoir solution, leading to larger single crystals (Chayen, 1997).

#### *d. Lysine methylation*

Reductive methylation of lysine residues was performed, by chemically modifying the primary amines (Lys and N-terminal residues) to tertiary amines. The positively charged Lys residues, lying on the surface of the protein, were transformed to small hydrophobic residues with the addition of a trimethyl group, in order to reduce the surface entropy of the protein and to increase the probability of crystal growth (Rayment, 1997; Walter et al., 2006). For this purpose, 1 mg of DSCAM Ig1-Ig8 GnTI after SEC (buffer described at [Table M-9](#)) was added to 20  $\mu$ l of freshly prepared 1 M dimethylamine-borane complex (ABC) in 40  $\mu$ l 1M formaldehyde. The mixture was incubated at 4 °C for 2 h under gentle shaking. The process was repeated, by adding 20  $\mu$ l ABC and 40  $\mu$ l formaldehyde and the mixture was further incubated for 2 h. A final aliquot of 10  $\mu$ l ABC was added and the reaction continued over night at 4 °C. The reaction was quenched with the addition of 20 mM Tris-HCl pH 7.5 and the methylated protein was purified with SEC using 20 mM Tris-HCl pH 7.5 in 150 mM NaCl. The MW of the methylated sample was assessed by MS.

#### *e. Endoglycosidase treatment*

Glycosylation is a post-translational modification of proteins, involved in diverse biological functions such as signaling processes, cellular trafficking, protein folding and stability. The flexibility of the different types of glycans lying on the vicinity of proteins residues, their size and their charge, in most cases introduce an increased level of conformational and chemical heterogeneity that might prevent protein crystallization (Aricescu and Owens, 2013). Protein deglycosylation is often used as a tool in order to decrease this heterogeneity. Endoglycosidase H (EndoH) was used

in this study to remove high mannose *N*-glycans from DSCAM Ig1-Ig8 when expressed in HEK 293S cells (GnTI<sup>-</sup>) (§2.2.4.1) in order to improve the quality of the diffracting crystals. The activity of the enzyme depends on the type of the glycosylation and it is only specific for a subset of glycans. Therefore, EndoH does not have the ability to cleave complex glycans, which are usually present in HEK 293T cell line (Aricescu et al., 2006; Freeze and Kranz, 2010). More specifically, it cleaves asparagine-*N*-linked high mannose oligosaccharides and some hybrid-type *N*-glycans (Figure M-3), leaving one GlcNAc residue attached to the asparagine site (Chang et al., 2007; Freeze and Kranz, 2010).

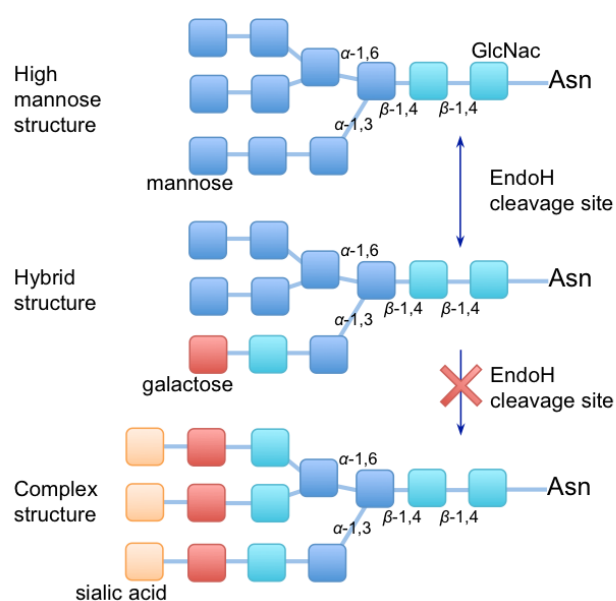


Figure M-3. Endoglycosidase H cleavage sites

10-200  $\mu\text{g}$  of DSCAM Ig1-Ig8 GnTI<sup>-</sup> were treated with 2  $\mu\text{l}$  EndoH, 37.5  $\mu\text{l}$  deionized water, 10  $\mu\text{l}$  of Reaction Buffer 1 in 2 % SDS, in the presence of 150 mM  $\beta$ -mercapthoethanol (total reaction volume 100  $\mu\text{l}$ ) (Table M-17). The mixture was incubated at RT for 24 h and samples were taken at different time intervals. The protocol was also repeated using Reaction Buffer 2. The activity of EndoH was assessed by monitoring the reduction of the MW of DSCAM by SDS-PAGE and MS was used to verify the deglycosylation of the protein sample.

Table M-17. Deglycosylation of DSCAM Ig1-Ig8 GnTI<sup>-</sup> with EndoH.

Components	Concentration	Volume ( $\mu$ l)
EndoH	1.0 unit* in 20 mM Tris-HCl pH 7.5, 25 mM NaCl	2
Reaction Buffer 1	250 mM NaH <sub>2</sub> PO <sub>4</sub> pH 5.5	10
Reaction Buffer 2	50 mM MES pH 6.0 150 mM NaCl	10
MilliQ H <sub>2</sub> O		37.5
Glycoprotein	10-200 $\mu$ g	Vt= 100

\* One unit will release the N-linked oligosaccharides from 60  $\mu$ moles of ribonuclease B per hour at 37 °C (pH 5.5).

#### f. Crystallographic data collection

Protein crystals are a periodic assembly of protein molecules in an ordered array. In order to determine their structure, crystals are exposed to a monochromatic X-ray beam using either in-house or synchrotron radiation sources. Part of the X-rays that impinge on the sample, are scattered from the electron clouds of the atoms that compose the protein and the resulting diffraction pattern is recorded by a detector. The diffraction pattern is formed by spots (reflections) that reveal the position and the intensity of each reflection. In a process called indexing, the position of each spot in one or more diffraction images (described by Miller indices h, k, l) is used to determine the crystal lattice parameters and its orientation. The intensities of all reflections for each image are integrated and the values derived by different images are scaled. A list of reflections along with their intensities is produced and is further used to calculate the structure factors i.e. the resultant amplitude and phase of scattering of all the electron density distribution of one unit cell. The structure factor of all reflections  $\mathbf{F}(hkl)$  is described from the following equation:

$$\mathbf{F}(hkl) = F(hkl) \exp i\alpha(hkl) = \sum_j f_j \exp [2\pi i (hx_j + ky_j + lz_j)]$$

\*  $i$ , is the imaginary number

\*  $\alpha(hkl)$  is the phase of the diffracted wave

\*  $F(hkl)$  is the amplitude of the wave

\*  $f_j$  is the scattering factor of the  $j^{\text{th}}$  atom with positional coordinates  $x_j, y_j, z_j$ . The summation is over all atoms of the unit cell

In a diffraction experiment where the phase cannot be measured directly, the structure is solved either by molecular replacement (MR) or by experimental phasing (single or multiple-wavelength anomalous dispersion (SAD/MAD), single or multiple

isomorphous replacement (SIR/MIR) and their combination: single or multiple isomorphous replacement with anomalous scattering (SIRAS/MIRAS)) and the Fourier transformation is used to calculate the electron density map ( $\rho(x,y,z)$ ), using the following formula:

$$\rho(x,y,z) = V^{-1} \sum_{hkl} \mathbf{F}(hkl) \exp(-2\pi i (hx + ky + lz))$$

\* V, is the volume of the unit cell

In the case where MR is used, initial phases are derived from a known structure of a homologous protein, which can be used as a search model. The model is considered as a rigid body and is subjected to rotation and translation within the unit cell. The six-dimensional transformation comprising a set of Euler angles and a matrix for translation is applied to the model. Model building, followed by alternate cycles of restrained refinement against experimental data and manual corrections to the atomic coordinates after visual inspection of the electron density maps, resulted in the final 3D structure of the target protein. In experimental phasing, where no previous model is available, a heavy atom is introduced into the protein crystal lattice either by engineering the amino acid chain directly (e.g. selenomethionine derivative) or by soaking or co-crystallizing the protein crystals with a heavy atom compound. The phase problem can then be solved with one or more heavy atom derivatives due to the increased contribution of the heavy atom to the overall scattering intensity (Blundell and Johnson, 1976; Rhodes, 2000). Alternatively, atoms such as S, Br or metals already present in the proteins could be exploited to obtain initial phases using SAD or MAD methods.

X-ray diffraction data for DSCAM Ig1-Ig9, DSCAM Ig1-Ig8 and DSCAM Ig1-Ig4, were collected at PETRA III MX beamlines (P13 and P14, EMBL-Hamburg, DESY, Germany). P13 is a tunable beamline (4.5 - 17.5 keV) that offers beam conditions with a beam-size of 30 x 20  $\mu\text{m}^2$ . P14 is also a tunable beamline (6 - 20 keV) that offers the option for micro-beam conditions (5 x 5  $\mu\text{m}^2$ ) and has a beam-divergence below 0.3 mrad, for data collection. One advantage of P14 is that focusing and de-focusing beam conditions can be easily adjusted and controlled by the software interface in order to work with different beam sizes and profiles. Both are high-brilliance beamlines with a flux in the range of  $10^{13}$  photons/sec. The ID29 beam line, dedicated to native and anomalous dispersion collection with an energy range of 6 - 20 keV (ESRF, Grenoble, France) was also used to collect native DSCAM Ig1-Ig4

diffraction data. All beam lines are equipped with a PILATUS 6M detector (Dectris, Switzerland) enabling shutter-less oscillation data collection. For all protein constructs, crystals were carefully handled at the same temperature that they were grown, were mounted on cryo-loops (Table S1) and they were flash-cooled in liquid nitrogen for cryo-data collection (100 K). Prior to flash-cooling, cryo-protectant solutions with different concentrations of either glycerol, ethylene glycol, polyethylene glycol 400 (PEG 400) or isopropanol, were prepared based on the composition of the mother liquor of the crystallization conditions, in order to prevent ice formation by the nitrogen cryo-stream, during data collection, that would destroy the crystal lattice.

For DSCAM Ig1-Ig4 both MR and experimental phasing were attempted. For experimental phasing, a range of heavy atoms was screened in order to examine their binding ability and the presence of the anomalous signal in the data collected. Protein crystals were soaked into fresh heavy atom solutions that contained complexes of  $\text{Pb}^{+2}$ :  $\text{Pb}(\text{CH}_3\text{COO})_2 \cdot 3\text{H}_2\text{O}$ ,  $\text{Pt}^{+2}$ :  $\text{K}_2\text{PtCl}_4$ ,  $\text{Yb}^{+3}$ :  $\text{YbCl}_3 \cdot \text{H}_2\text{O}$  and  $\text{Gd}^{3+}$ :  $\text{Gd-HPDO3A}$ . More specifically a DSCAM Ig1-Ig4 WT crystal was transferred to 10  $\mu\text{l}$  of a solution, containing the mother liquor reagents and the heavy atom in a final concentration of 5 - 100 mM in various soaking times (10 min - 24 h). Crystal quality was monitored with the aid of a microscope for possible signs of crystal cracking or dissolving effects. The remaining heavy atom solution, which surrounded the crystal, was removed by soaking the crystals in a reservoir solution containing 20-25 % PEG 400 or ethylene glycol as a cryoprotectant. Crystals were then exposed to X-rays and data were collected at the absorption edge of each heavy atom (Table M-18).

Table M-18. X-ray absorption edges.

Heavy Atom	Edge L-III (keV)	Å
Pb	13.0352	0.9511
Pt	11.5637	1.0722
Yb	8.9436	1.3863
Gd	7.2428	1.7118

*g. Data processing, phasing and refinement*

Data collection at synchrotron radiation sources was performed using mxCuBE v2 (Gabadinho et al., 2010) and the diffraction images were displayed with Adxv software (Arvai Andrew, Scripps Research Institute). The images were indexed either with MOSFLM (Leslie et al., 2007) or with XDS (Kabsch, 2010). XDS was also used to integrate the reflections, followed by POINTLESS (Evans, 2006) to determine the crystal lattice and symmetry. The integrated reflections were further scaled and merged using SCALA or AIMLESS (Evans, 2006). For MR, either MOLREP (Vagin and Teplyakov, 1997) or PHASER (McCoy et al., 2007) was used and for experimental phasing, the SHELX package (Pape and Schneider, 2004; Sheldrick, 2008). Refinement of the search model against experimental data was carried out with REFMAC (Murshudov et al., 1997, Murshudov et al., 2011). The aforementioned modules (except for XDS) are implemented in CCP4 suite (Winn et al., 2011). Visual inspection of the electron density maps and model building were conducted with COOT (Emsley and Cowtan, 2004). Schematic representation of protein molecules and graphics were prepared with UCSF CHIMERA Package (Pettersen et al., 2004).





## Chapter 3

### RESULTS AND DISCUSSION

#### 3.1 Results

Down Syndrome Cell Adhesion Molecule (DSCAM) is a cell surface receptor and its ectodomain consists of ten Ig-like C2 and six FNIII domains. It is expressed in neurons during development and is involved in axon guidance, cell aggregation and cell adhesion during dendrite arborization (Kise and Schmucker, 2013; Zinn, 2007). Previous studies in *D. melanogaster* suggest that Dscam in insects has a dual role; it is implicated mainly in self-recognition among dendrites arising from the same neuron cell and also in heterophilic binding with pathogen molecules (Watson et al., 2005; Schmucker and Chen, 2009). Dscam *trans* homophilic binding between opposed dendrites, generates a repulsive signal not allowing them to establish connections. This process is called self-avoidance and prevents undesirable interactions between sister dendrites. This discrimination among dendrites of the same cell, seems to be achieved due to the high diversity that Dscam exhibits in *Drosophila* (~38,000 isoforms with ~19,000 ectodomains) and studies have shown that it is the specificity of certain domains that is responsible for the homophilic interactions and self-avoidance (Wojtowicz et al., 2007; Sawaya et al., 2008) (Figure I-9).

On the contrary, DSCAM in vertebrates does not exhibit molecular diversity, resulting in an alternative role of DSCAM in mammals. Experiments in mouse retina (Fuerst et al., 2008), have shown that although DSCAM is involved in self-avoidance through homophilic binding interactions, it does not appear to be associated with self-discrimination processes, as in insects. DSCAM is additionally found to be implicated in neuronal development and function (Yamakawa et al., 1998) and upon expression in commissural axons it is involved in axon guidance through heterophilic binding with guidance cues like netrin-1 and other receptors like DCC (Liu et al., 2009; Ly et al., 2008). Netrin-1 is a glycosylated protein involved in cell signaling in the developing and adult nervous system and it has been characterized as a guidance cue that can act either from a long or a short distance, mediating attraction or repulsion depending on its binding partners. It can also alter the formation of different tissues by mediating cell migration, cell-cell interactions and cell-extracellular matrix

adhesion (Lai Wing Sun et al., 2011). Upon binding to DSCAM, netrin mediates attraction and promotes axonal growth.

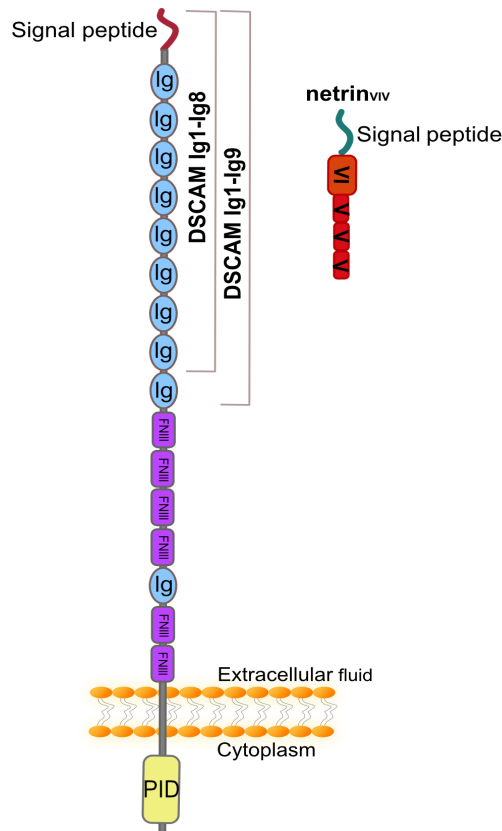
It is anticipated that characterization of human DSCAM and of DSCAM in complex with netrin-1, will assist the efforts of unraveling the binding mechanism in vertebrates and identify the differences upon DSCAM homophilic and heterophilic interactions compared to the same mechanism in insects. To this end, structural and biophysical experiments were performed with the aforementioned proteins and are described in this section.

### ***DSCAM Ig1-Ig8, DSCAM Ig1-Ig9, DSCAM Ig1-Ig9:netrin<sub>VIV</sub>***

#### **3.1.1 Construct design rationale of DSCAM Ig1-Ig8 & Ig1-Ig9**

According to previous structural studies of Dscam Ig1-Ig8 in *D. melanogaster*, the Ig domains implicated in the homodimerization, are Ig2, Ig3 and Ig5-Ig7 (Meijers et al., 2007; Sawaya et al., 2008), out of which domains Ig2, Ig3 and Ig7 were shown to be essential for isoform specificity. This specificity is important for homophilic recognition and binding interactions among the different isoforms arising from opponent dendrites (Wojtowicz et al., 2004, 2007). With the aim to explore DSCAM homodimerization in humans and assess the differences between the two species, the same construct (DSCAM Ig1-Ig8) was initially prepared (Figure R-1).

Previous studies performed by Tessier-Lavigne and co-workers, suggested that the DSCAM binding interface with netrin-1 possibly lies between domains Ig7-Ig9 (Ly et al., 2008). In order to further investigate DSCAM heterophilic interactions in the presence of netrin-1, an Ig domain was added to the existing DSCAM Ig1-Ig8 construct corresponding to the 9<sup>th</sup> Ig extracellular domain (Figure R-1). More specifically, the research focused on exploring the interactions of DSCAM with netrin-1 upon complex formation and investigating whether DSCAM Ig1-Ig9 homodimerization is abolished upon binding to netrin. Netrin-1 (*Homo sapiens* UniProtKB-O95631) is composed of a laminin domain VI, which is globular, followed by three epidermal growth factor (EGF-like) repeats of the  $\gamma$ -chain of laminins V (EGFV-1, EGFV-2, EGFV-3). These N-terminal domains are linked to a C-terminal netrin-like module (NTR) (Rajasekharan and Kennedy, 2009). The netrin-1 construct used in this study (netrin<sub>VIV</sub>), consists of one laminin domain (VI) and three laminin EGF-like domains (V) (Finci et al., 2014) (Figure R-1).



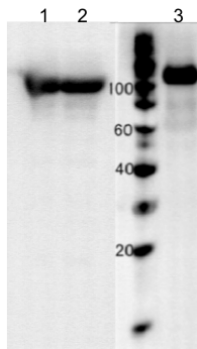
**Figure R-1.** Schematic representation of protein constructs for DSCAM and netrin-1. DSCAM extracellular domain comprises ten *N*-terminal Immunoglobulin-like domains (Ig-like C2-set) and six Fibronectin type III (FNIII) domains. DSCAM Ig1-Ig8 and DSCAM Ig1-Ig9 constructs were designed consisting of eight and nine Ig-like domains, respectively. Netrin<sub>VIV</sub> comprises one laminin domain (VI) and three EGF-like domains (V). The secretion signal peptide used for DSCAM expression was the native one. For netrin<sub>VIV</sub> the signal peptide used was from Human Pregnancy Specific Beta-1-Glycoprotein (PSG-1), a protein expressed in high yield in HEK cells.

### 3.1.2 Purification & biophysical characterization of DSCAM Ig1-Ig8 & Ig1-Ig9

The DNA sequence encoding the *N*-terminal eight Ig domains of DSCAM (*Homo sapiens* UniProtKB-O60469) (DSCAM Ig1-Ig8) was obtained as a synthetic gene by GenScript. For DSCAM Ig1-Ig9, overlap extension cloning was performed between DSCAM Ig1-Ig8 and DSCAM Ig9 domain (obtained by GenScript) (§2.2.3.3) with the primers listed in [Table S5](#). Both genes were recloned into the pXLG vector, equipped with a *C*-terminal His<sub>6</sub>-tag, using DSCAM native secretion signal. After plasmid amplification in DH5a cells (*E. coli*), the DNA of both constructs was introduced into HEK 293T and HEK 293S cells (devoid of GnTI) and proteins were secreted to one Liter of culture media. The protein expression was assessed by WB ([Figure R-2](#)).

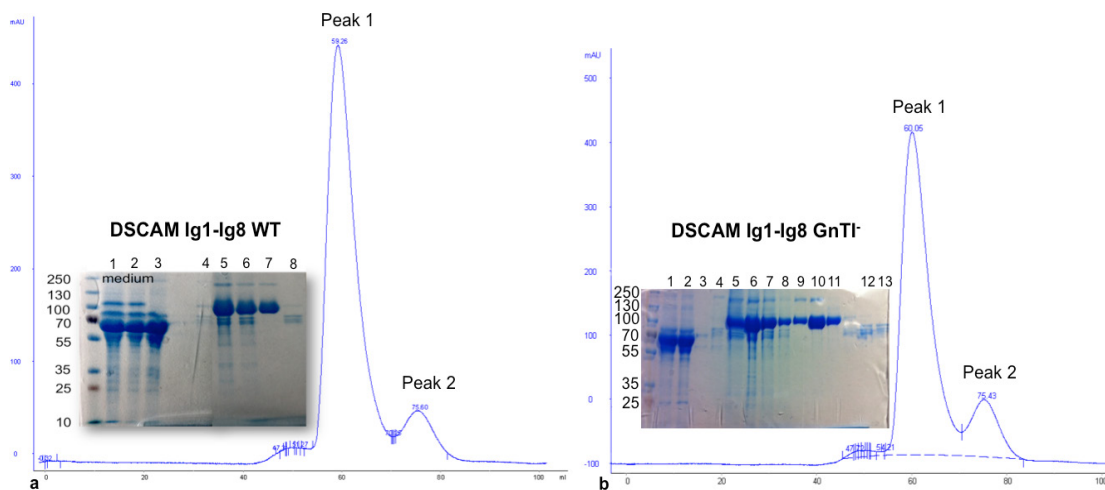
Both proteins were purified with affinity chromatography, followed by SEC (§2.2.5). One Liter of cell culture resulted in ~10 mg for DSCAM Ig1-Ig8 WT and in ~8

mg for DSCAM Ig1-Ig8 GnTI<sup>-</sup>, after the final purification stage. Similarly, DSCAM Ig1-Ig9 resulted in ~30 mg and ~20 mg for the WT and the GnTI<sup>-</sup> variant, respectively.



The purification profile of DSCAM Ig1-Ig8 (both for WT and GnTI<sup>-</sup>), after SEC, indicated that the elution peaks corresponded to different oligomeric states. The protein purification was evaluated by SDS-PAGE (Figure R-3).

**FigureR-2.** Western blot from large-scale expression (1 Liter cell culture) of DSCAM Ig1-Ig8 (lanes 1,2) and DSCAM Ig1-Ig9 (lane 3) in HEK 293T cell line.



**Figure R-3.** SEC profile of DSCAM Ig1-Ig8 WT (a) and GnTI<sup>-</sup> (b) and SDS-PAGE after IMAC and SEC purifications. **a.** lanes 1-3: protein in the medium before purification, lane 4: washing step after IMAC, lanes 5,6: eluates after IMAC, lane 7: eluate corresponding to peak 1 after SEC, lane 8: eluate corresponding to peak 2 after SEC. **b.** lanes 1,2: protein in the medium before purification, lanes 3,4: washing steps after IMAC, lanes 5-8: eluates after IMAC, lanes 9-11: eluates corresponding to peak 1 after SEC, lanes 12,13: eluates corresponding to Peak 2 after SEC.

In order to assess the oligomeric state of DSCAM Ig1-Ig8 after purification, further experiments were performed with FPLC coupled with static light scattering (SLS), using Malvern's OmniSEC GPC/SEC multidetector (§2.2.7.3). The sample was filtered, loaded on a Superdex 200 10/300 GL column and was eluted with a flow rate of 0.4 ml/min. The first peak corresponded to dimeric DSCAM Ig1-Ig8 WT, with a

determined MW at  $180 \pm 10$  kDa. The MW was calculated by UV measurements derived from baseline corrected RALS intensities and also by comparison to a protein that was used as a standard with an established MW (BSA). The second elution peak corresponded probably to monomeric protein, with an estimated MW around  $98 \pm 25$  kDa, due to limitations in resolution derived by the SEC column used (Figure R-4).

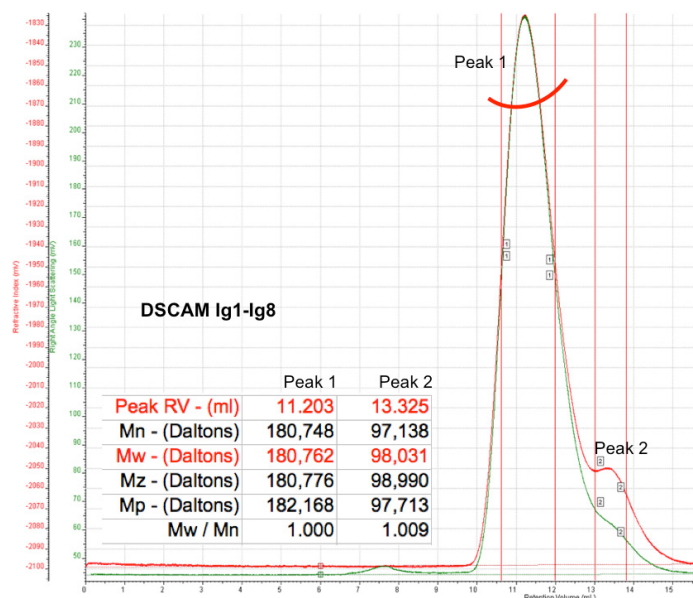
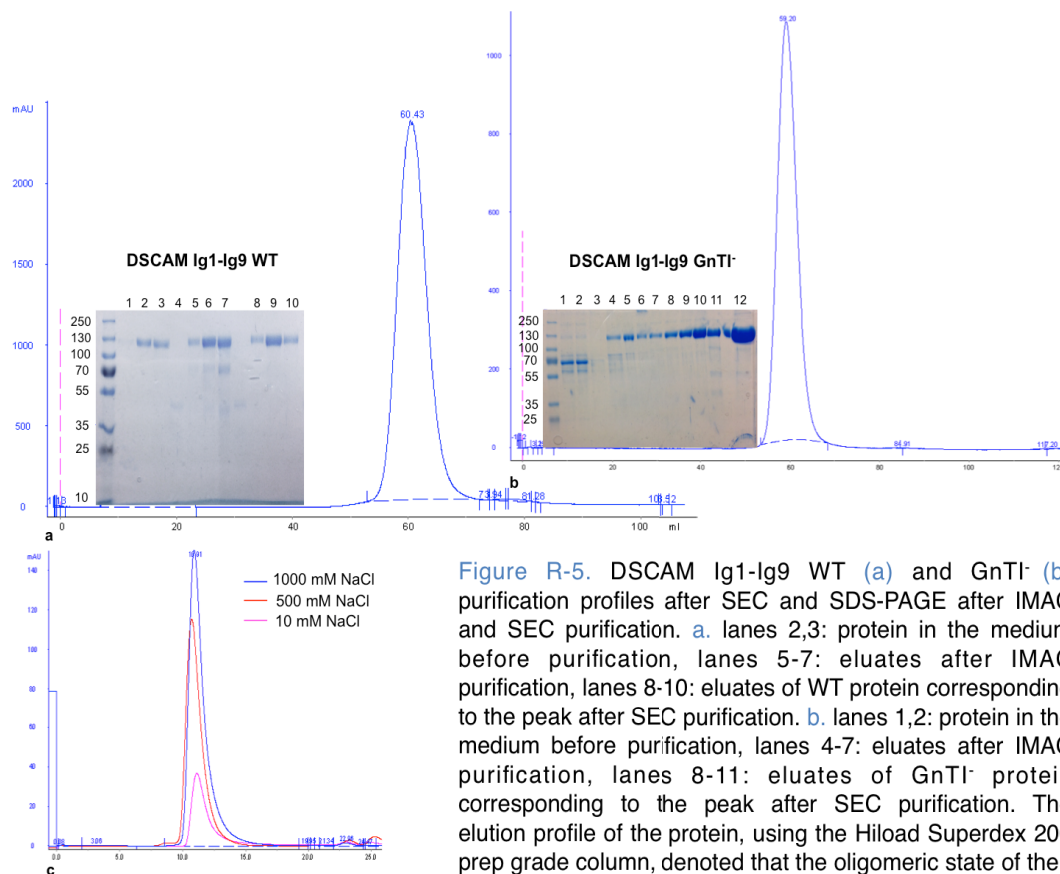


Figure R-4. DSCAM Ig1-Ig8 WT oligomeric state analysis with FPLC coupled with SLS detectors, using Malvern's OmniSEC GPC/SEC system. The protein in solution exists in two states with the dimer being the predominant species (Peak 1) with a MW determination of  $180 \pm 10$  kDa. Peak 2 corresponds to monomeric DSCAM with MW estimation to  $98 \pm 25$  kDa. The red line corresponds to the refractive index (RI) and the green line to RALS.

On the contrary, the purification profile of DSCAM Ig1-Ig9 showed that it was eluted only in one oligomeric state (as a dimer). This result suggested that the addition of Ig9 domain further stabilized the homodimeric interactions of the protein, compared to DSCAM Ig1-Ig8. Its oligomeric state was independent of salt concentration as it was shown by SEC purification profiles examined under various salt concentrations (10 mM - 1M NaCl) (Figure R-5c). The purity of the protein samples after IMAC and SEC was assessed by SDS-PAGE for both WT and GnTI<sup>-</sup> and the results are shown in Figure R-5a,b. The molecular weight of the protein was  $\sim 114$  kDa for the WT and  $\sim 109$  kDa for the GnTI<sup>-</sup> as it was evaluated by MS analysis (the MW corresponded to monomeric DSCAM measured under denaturing conditions).



**Figure R-5.** DSCAM Ig1-Ig9 WT (a) and GnT1<sup>-</sup> (b) purification profiles after SEC and SDS-PAGE after IMAC and SEC purification. a. lanes 2,3: protein in the medium before purification, lanes 5-7: eluates after IMAC purification, lanes 8-10: eluates of WT protein corresponding to the peak after SEC purification. b. lanes 1,2: protein in the medium before purification, lanes 4-7: eluates after IMAC purification, lanes 8-11: eluates of GnT1<sup>-</sup> protein corresponding to the peak after SEC purification. The elution profile of the protein, using the Hiload Superdex 200 prep grade column, denoted that the oligomeric state of the protein was dimeric. c. SEC purification of DSCAM Ig1-Ig9 WT in 50 mM MES pH 6.0 using 10mM of NaCl (in pink), 500 mM of NaCl (in red) and 1M NaCl (in blue). Superposition of elution profiles showed that the oligomeric state of the protein was independent of salt concentration.

protein was dimeric. c. SEC purification of DSCAM Ig1-Ig9 WT in 50 mM MES pH 6.0 using 10mM of NaCl (in pink), 500 mM of NaCl (in red) and 1M NaCl (in blue). Superposition of elution profiles showed that the oligomeric state of the protein was independent of salt concentration.

Thermofluor was also performed to examine the stability of DSCAM. A number of conditions were tested using different buffer solutions, salts and ligands. The results demonstrated that DSCAM Ig1-Ig8, in the presence of some of the conditions (e.g. 3 % ethylene glycol, 100 mM CH<sub>3</sub>CO<sub>2</sub>K, 5 % PEG 400), exhibited a second unfolding event (Figure R-6). This observation indicated that DSCAM Ig1-Ig8 presented an intermediate substructure upon unfolding. Initially it was thought that this substructure could be either derived by a more rigid domain within the monomer or by dissociation of DSCAM dimer. When those conditions, which indicated a more stable protein population, were tested in the purification process (e.g. 150mM NaCl, 2 mM DTT, 5 % glycerol) the eluted sample showed that DSCAM Ig1-Ig8 was still a mixture of monomers and dimers, with the dimeric state being the prevalent one.

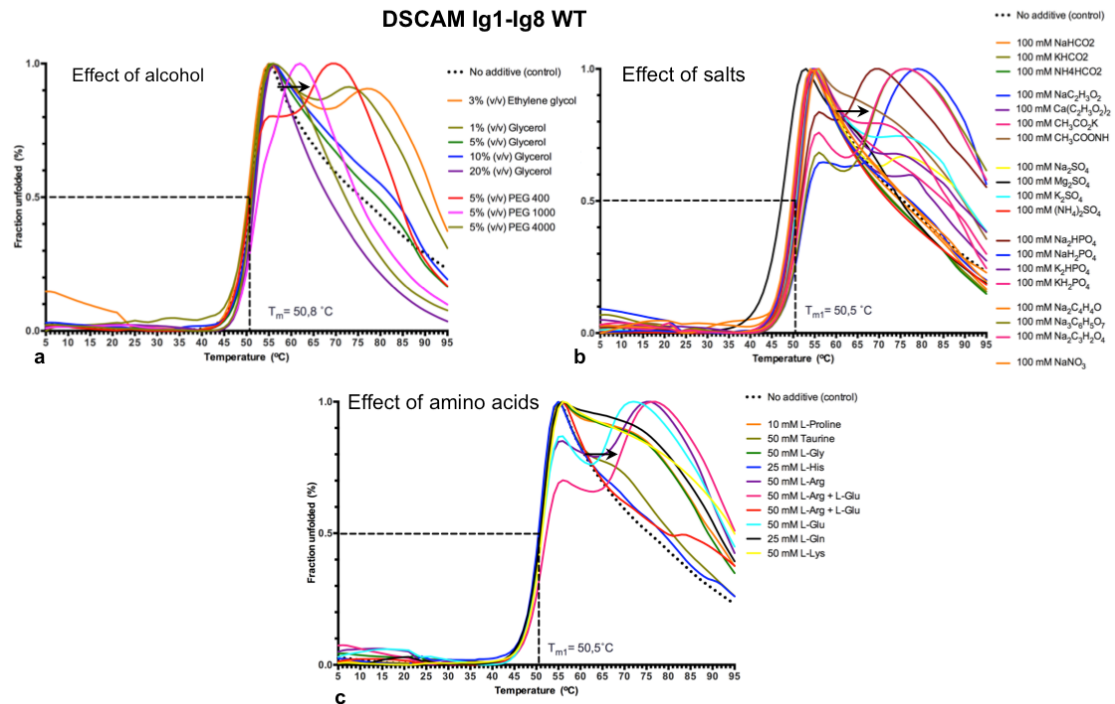
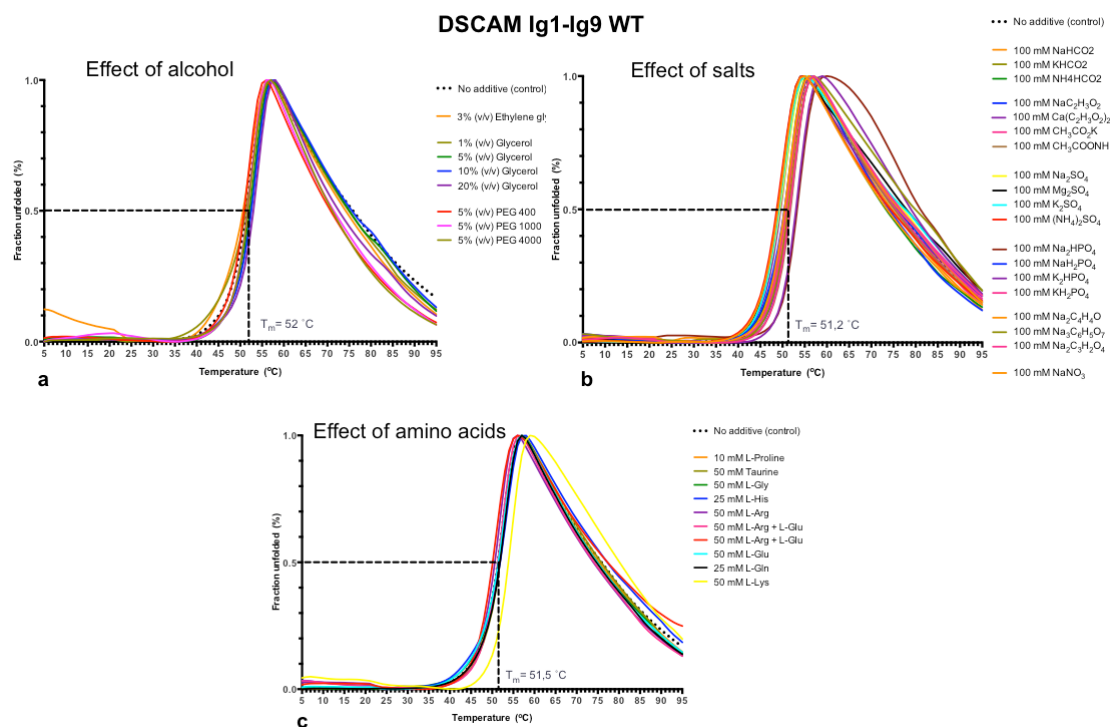


Figure R-6. Selected graphs of thermofluor assays for DSCAM Ig1-Ig8 WT, examined under various conditions, indicated a melting point of  $\sim 50$  °C. A second unfolding event was observed (marked with an arrow) in the presence of conditions such as 3 % ethylene glycol, 100 mM  $\text{CH}_3\text{CO}_2\text{K}$ , 5 % PEG 400, 50 mM L-Glu, 50 mM L-Glu & L-Arg.

For DSCAM Ig1-Ig9, the corresponding melting curves obtained by thermofluor, suggested that the protein remained stable in the presence of all conditions tested, without exhibiting additional unfolding curves. This was in accordance with the purification results for DSCAM Ig1-Ig9, which showed that the protein was always eluted as a stable dimer. Therefore, the differences observed between the two constructs, might be attributed to the addition of domain Ig9 that seemed to stabilize the dimer interactions of DSCAM Ig1-Ig9. Selected graphs are presented in Figure R-7.



**Figure R-7.** Selected graphs of thermofluor assays for DSCAM Ig1-Ig9 WT, examined under various conditions, indicated a melting point of  $\sim 52\text{ }^{\circ}\text{C}$ .

### 3.1.3 Structural characterization of DSCAM Ig1-Ig8 & Ig1-Ig9

#### 3.1.3.1 Small-angle X-ray scattering of DSCAM Ig1-Ig8 & Ig1-Ig9

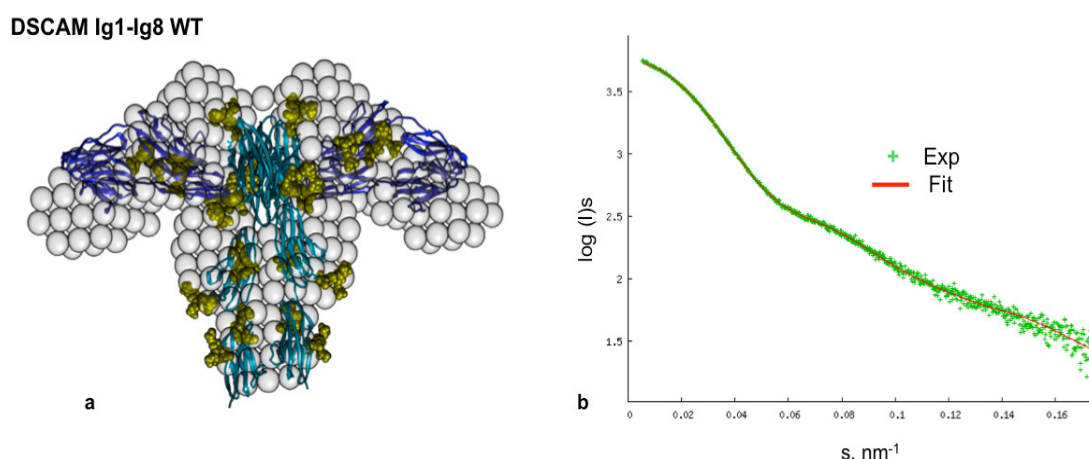
In order to shed light on DSCAM structural features and determine its overall shape, small-angle X-ray scattering experiments were performed with both DSCAM constructs (Ig1-Ig8 and Ig1-Ig9 in WT and GnTI<sup>+</sup> form). Data were collected at P12 (PETRA III, EMBL-Hamburg, DESY, Germany) in collaboration with Dr. Gundolf Schenk and Dr. Haydyn Mertens (Svergun Group). Regarding DSCAM Ig1-Ig8, although it was in equilibrium of dimers and monomers, shifted towards the dimeric state, (as it was observed after its purification and biophysical experiments, [Figure R-4](#)), measurements performed and reliable models based on the experimental data were only generated for the dimeric state after SEC ([Figure R-3: sample corresponding only to dimer population](#)). For DSCAM Ig1-Ig9, oligomerization state analysis demonstrated that it was dimer in solution (monodisperse population). A series of protein concentrations ([Tables S6-S7](#)) were tested and the extrapolated intensity at zero scattering angle  $I(0)$ , the radius of gyration ( $R_g$ ) and the maximum



interatomic distance ( $D_{\max}$ ) calculated for this concentration range, showed no concentration-dependent oligomerization for both constructs examined.

Two independent SAXS modeling approaches were employed: *ab initio* reconstruction directly from the data and rigid body refinement. The models were reconstructed by applying P1 (for an unbiased model reconstruction) and P2 symmetry. The theoretical scattering curves calculated from the generated models for both Ig1-Ig8 and Ig1-Ig9, using the two methods, were in good agreement with the experimental data obtained from SAXS measurements, with the calculated discrepancy value, which represents the quality of the fit,  $\chi^2$ , being at 1.0.

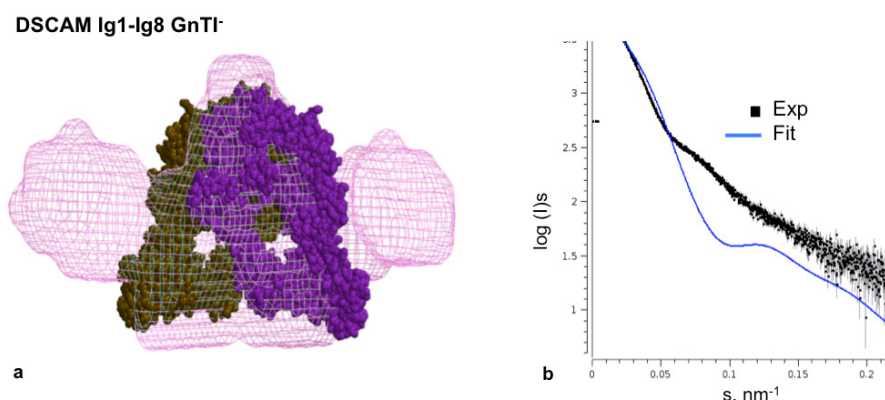
For DSCAM Ig1-Ig8, rigid body refinement was performed by defining the *N*-terminal Ig1-Ig4 domains as a rigid core, resulting in similar models with normalized spatial discrepancy (NSD) of 1.0 (measure of quantitative similarity). Superposition of DSCAM Ig1-Ig8 *ab initio* model with the model generated by the rigid body approach, demonstrated the agreement between the models obtained independently by the two methods used (NSD=1.1) (Figure R-8).



**Figure R-8.** SAXS analysis and model generation of DSCAM Ig1-Ig8 WT. **a.** Superposition of the *ab initio* model (transparent beads) with the model generated using the rigid body approach (domains Ig1-Ig4 are shown in dark blue & domains Ig5-Ig8 in light blue), depicted high similarity for the shape of DSCAM with NSD at 1.1. The theoretical *N*-glycosylation sites on DSCAM Ig1-Ig8 model (in yellow) were built using the GlyProt server (Bohne-Lang and Von der Lieth, 2005) based on DSCAM amino acid sequence. **b.** Calculated scattering curve of the *ab initio* model (in red) fits well the experimental data (in green) and is plotted as the logarithm of scattered intensity against the momentum transfer, with the quality of fit  $\chi^2$  being at 1.0 (Table R-1).

In order to test the working hypothesis, the *ab initio* model was compared to Dscam Ig1-Ig8 homologous crystallographic structure from *Drosophila*, with 34 % sequence identity (over 263 out of 780 amino acids) at 4.2 Å resolution (PDB entry

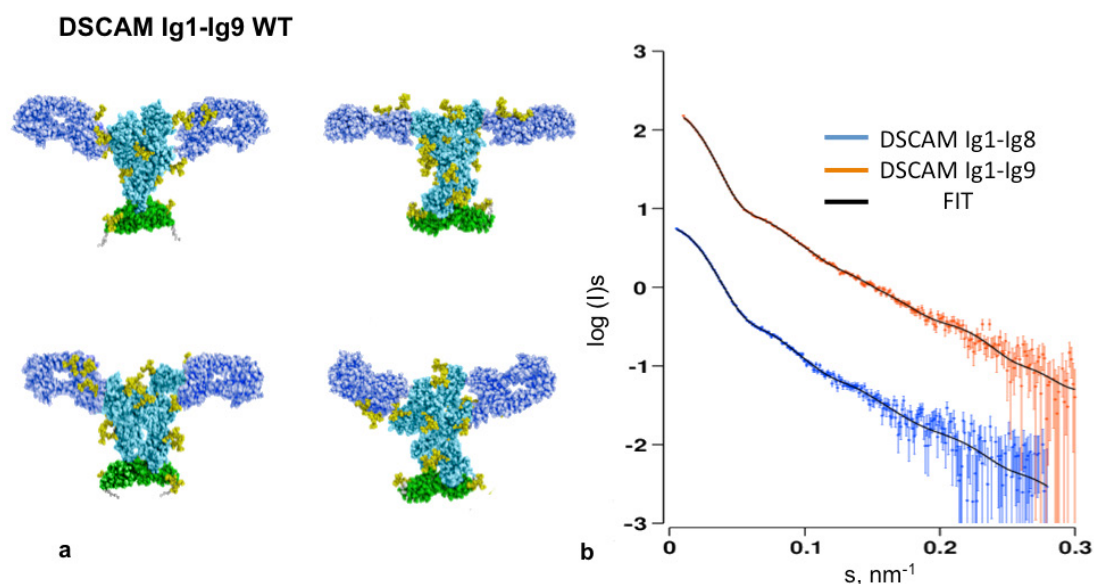
3DMK) (Sawaya et al., 2008). The results showed that the theoretical scattering curve, computed from the crystallographic structure, did not fit the experimental SAXS data for DSCAM ( $\chi^2 \sim 4.0$ ). DSCAM dimer in humans was more flexible and had differences from the *Drosophila* Dscam structure, mainly regarding domains Ig1-Ig4 (Figure R-9).



**Figure R-9.** DSCAM Ig1-Ig8 GnTI SAXS experimental data in solution in comparison to Dscam crystallographic structure from *Drosophila* (PDB entry 3DMK). **a.** Superposition of the *ab initio* SAXS envelope of DSCAM (shown in pink) to the crystallographic structure of Dscam Ig1-Ig8 (each monomer is represented in brown and purple) (Sawaya et al., 2008). **b.** The theoretical scattering curve (shown in blue) computed from the crystal structure in *Drosophila* with CRY SOL did not fit the experimental scattering data (shown in black). Data are plotted as the logarithm of scattered intensity against the momentum transfer ( $s$ ), with  $\chi^2$  values being above 4.0.

For DSCAM Ig1-Ig9, which was a dimer in solution, rigid body models were computed by restraining the *N*-terminal Ig1-Ig4 domains, either as a single rigid body using data obtained from DSCAM Ig1-Ig4 SAXS experiments (a detailed analysis for DSCAM Ig1-Ig4 construct is presented in §3.1.8.1), or using a global refinement procedure using the program CORAL (Petoukhov et al., 2012), incorporating data collected from all DSCAM constructs, simultaneously, i.e. Ig1-Ig4, Ig1-Ig8 and Ig1-Ig9. Ten independent runs were conducted and similarity was assessed using the program DAMAVER (Volkov and Svergun, 2003), which also identified the most representative member of the ensemble. Since domain Ig9 appeared to stabilize the dimer interactions (based on purification results - Figure R-5) the models were constructed by defining this domain as a contact point, based on the  $C_\alpha$ - $C_\alpha$  distances for constraining the rigid body modeling. The results demonstrated that all dimer models shared common structural features assembling into a T-shape configuration. Furthermore, in every model generated, domains Ig1-Ig4 (Figure R-10a in dark blue)

were shown to be located at the periphery of this T-shaped molecule away from the dimeric interface. This approach resulted in an ensemble of models that fit equally well the experimental data with  $\chi^2$  values of 1.1 (Figure R-10b).



**Figure R-10.** a. Cluster of DSCAM Ig1-Ig9 WT rigid body models obtained by SAXS (domains are shown in dark blue: for Ig1-Ig4, light blue: for Ig5-Ig8, green: for Ig9 and yellow: for glycans) reconstructed using simultaneously experimental data from DSCAM Ig1-Ig4, Ig1-Ig8 and Ig1-Ig9. In every model, domains Ig1-Ig4 were located away from the core of the T-shaped molecule (shown in dark blue). b. The calculated scattering curve of the models (in black) fits well the experimental scattering curve for DSCAM Ig1-Ig9 (shown in orange) with  $\chi^2$  values of 1.1. Data are plotted as the logarithm of scattered intensity against the momentum transfer ( $s$ ).

**Table R-1.** Table of selected statistics for DSCAM Ig1-Ig8 and DSCAM Ig1-Ig9 derived from SAXS data collection. The complete table of statistics can be found at the Appendix Section.

Protein	$R_g$ Guinier (nm)	$D_{max}$ (Å)	Volume (nm <sup>3</sup> )	MW <sup>1</sup> (kDa)	MW <sup>2</sup> (kDa)	MW <sup>3</sup> (Da)	Oligomeric state
DSCAM Ig1-Ig8	5.8±0.2*	20±0.5*	494±50*	167±17*	200±20*	88183	Dimer
DSCAM Ig1-Ig9	6.4±0.2	22.3±0.5	726±70	226±20	207±20	96890	Dimer

\* The values for DSCAM Ig1-Ig8 correspond to the dimeric state.

MW<sup>1</sup>: calculation based on the absolute scattering intensity  $I(0)$ .

MW<sup>2</sup>: calculation based on BSA (used as a standard).

MW<sup>3</sup>: calculation based on sequence (for the monomer without glycosylation).

### 3.1.3.2 Electron microscopy of DSCAM Ig1-Ig8 & Ig1-Ig9

With the aim to identify which Ig domains are involved in homodimer interactions, structural studies of DSCAM Ig1-Ig8 and Ig1-Ig9 (both in WT and GnTI<sup>-</sup>) were performed with negative staining electron microscopy.

EM studies on DSCAM Ig1-Ig8 were conducted in collaboration with the Electron Microscopy Core Facility (EMBL-Heidelberg). A variety of stains (i.e. uranyl formate, uranyl acetate and ammonium molybdate) in the presence of NaCl (at concentrations ranging from 100 to 500 mM) were tested for both dimeric and monomeric states after purification (Figure R-3, Peak 1 & Peak 2, respectively) corresponding to WT and GnTI<sup>-</sup> variant. The raw images obtained, showed a mixture of monomers, dimers and larger aggregates that coexisted in the sample. DSCAM samples in the presence of 500 mM NaCl exhibited less aggregates, compared to the ones in lower salt concentrations (Figure R-11). Despite the different conditions and stains examined, the enhanced heterogeneity observed in the population, prevented further particle analysis using this construct.

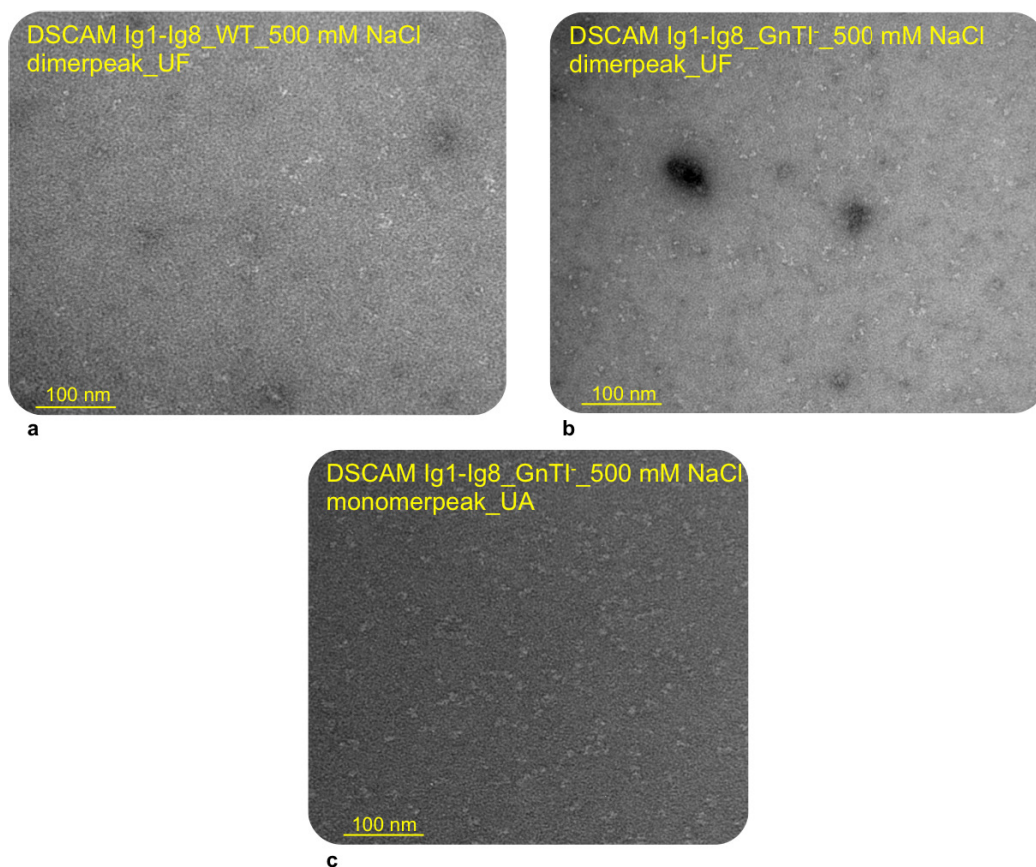


Figure R-11. DSCAM Ig1-Ig8 WT (a) and GnTI<sup>-</sup> (a,c) raw images after negative staining, stained with 0.75 % uranyl formate or uranyl acetate (using 50 mM Citrate pH 6.0 and 500 mM NaCl) corresponding to dimer (a,b) and monomer (c) populations after SEC purification. The images obtained demonstrated a mixture of monomers, dimers and larger aggregates.

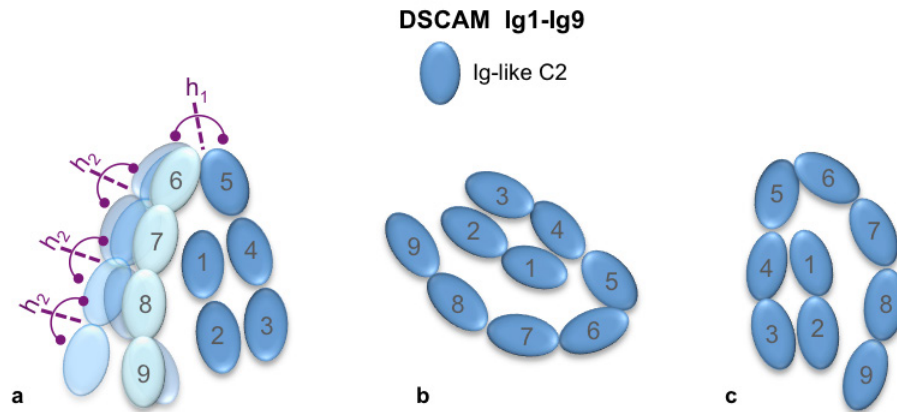
Electron microscopy studies on DSCAM Ig1-Ig9 were performed by Skiniotis Group (Life Sciences Institute, University of Michigan). Preliminary analysis of the raw images showed that most of the population was monomeric with only a few particles depicting a dimer of DSCAM. Based on these initial observations, reference-free alignment and classification was performed to obtain 2D averages of DSCAM Ig1-Ig9 (Figure R-12).

More specifically, single-particle analysis demonstrated that although DSCAM Ig1-Ig9 was a dimer in solution (in concentration range 0.3 - 10 mg/ml), most of the specimen population appeared to be in monomeric state. Approximately 13,000 particle projections of DSCAM Ig1-Ig9 WT were subjected to ISAC to address structural heterogeneity and assess the differences in the population between monomers and dimers. These resulted in 222 classes accounting for 9,965 particle projections from 20 rounds of classification. Part of the projections appeared only as single rods (Figure R-12 shown in green), some exhibited a partial monomer and others revealed the overall structure of DSCAM Ig1-Ig9 in different conformations and orientations (Figure R-12 shown in blue and red, respectively). The horseshoe-shaped arrangement (comprising four *N*-terminal Ig domains), which was observed in projections that disclosed the overall structure, seemed to be similar to that of Dscam Ig1-Ig8 in *D. melanogaster*. Thorough analysis of these projections revealed that domains Ig1-Ig5 exhibited a more stable assembly, whereas domains Ig6-Ig9 formed a linear-like arrangement, moving away or leaning towards the rigid horseshoe configuration i.e. projections 70 & 98 (Figure R-12 marked with a red rectangle). Moreover, a closer look on some projections, where the horseshoe maintained the same orientation, (e.g. 136 and 140 Figure R-12 marked with a red circle) indicated that additional degrees of freedom were probably introduced among domains Ig7 to Ig9, resulting in a number of different conformations.



**Figure R-12.** Single-particle analysis of DSCAM Ig1-Ig9 WT in 50 mM MES pH 6.0, 150 mM NaCl. Reference-free alignment of particle projections was performed with ISAC and resulted in 222 classes accounting for 9,965 particle projections after 20 rounds of classification. DSCAM Ig1-Ig9 is depicted as a monomer in different conformations and orientations and selected projections are marked in red. Projections marked with the same colour and shape show DSCAM molecule in which the horseshoe maintains the same orientation but the rest of the domains are flexible. Partial monomers of DSCAM are shown in blue and domains that appeared as single rods are shown in green. A partial dimer of DSCAM where only the first four Ig domains can be visualized for one partner, is shown in yellow.

The conformational flexibility of DSCAM could be attributed to a potential hinge located at any place in the tandem of Ig5-Ig9. The presence of hinges would explain the rotational freedom around an axis driving the last three or four Ig domains to move as a rigid body, either towards or away from the horseshoe arrangement (Figure R-13a).



**Figure R-13.** Schematic representation of DSCAM Ig1-Ig9 in different conformations (a) and orientations (b,c) after two-dimensional reference-free classification. a. The first four *N*-terminal Ig domains adopt a rigid horseshoe conformation (Ig1-Ig4 in dark blue) with domains Ig6-Ig9 forming a linear-like arrangement (in light blue), which moves either away or leaning towards the horseshoe assembly. The pronounced variations in conformations might be a result of potential hinges that join domains Ig5-Ig6 ( $h_1$ ) or can be located anywhere among domains Ig6-Ig9 ( $h_2$ ).

The domain structure of DSCAM (construct Ig1-Ig9) appears to follow the architecture of other cell adhesion molecules and its individual domains have an Ig-like folding, which is formed by antiparallel  $\beta$ -strands, organized in a double  $\beta$ -sheet. Sequence alignment of DSCAM Ig1-Ig9 with selected protein receptors bearing similar domain organization and hinge regions, was performed, in order to identify potential hinges in DSCAM and justify the observations made by EM (Figure R-14). Two hinge regions were identified in Dscam Ig1-Ig8 from *D. melanogaster* (PDB entry 3DMK) (Sawaya et al., 2008) and in Titin Ig65-Ig70 (PDB entry 3B43) (von Castelmur et al., 2008), between Ig domains 4-5-6 and between domains 65-66-67, respectively (Figure R-14 shown in orange and green rectangles, accordingly). These regions boost the conformational freedom of the main chain through specific linkers (i.e. residues ERK in titin, Figure R-14 shown in green rectangles) that act as bending points. In most of the proteins examined, the observed conformers of each moving substructure seemed to depend on the length of the linker that dictated whether it would function as a hinge, allowing modular chain motions. The areas between Ig domains that consisted of two residues or of zero-length linker were more tightly connected, allowing only limited flexibility of the individual domains and not a profound chain bending (von Castelmur et al., 2008).

In the case of Dscam, a linker, located between domains Ig4-Ig5, comprising five amino acids (i.e. residues GGRFD, Figure R-14 shown in orange rectangle) was

identified as a hinge region (using the DynDom server and by measuring the angle of the domains using the Dscam Ig1-Ig8 structure) allowing rotation of Ig1-Ig4 (as rigid-body) relative to the rest of the molecule (Ig5-Ig8) (Hayward and Berendsen, 1998; Sawaya et al., 2008). In addition, a linker (connecting domains Ig6-Ig7) consisting of five residues induced as well a conformational bending motion. The hinge mechanism occurring around the specific domains seemed to lead to interdomain flexibility, necessary for hosting the second molecule of Dscam during homodimerization.

Similarly, the presence of two hinge regions was identified in tyrosine kinase receptor KIT (PDB entry 2EC8) (Yuzawa et al., 2007) and in Colony-stimulating factor 1 (PDB entry 4LIQ) (Ries et al., 2014). These hinges were between domains Ig3-Ig4-Ig5 and domains Ig2-Ig3-Ig4, respectively (Figure R-14, shown in purple and yellow rectangles, accordingly) whereas the tyrosine kinase receptor Flt3 (PDB entry 3QS7) (Verstraete et al., 2011) had one hinge axis among domains Ig3-Ig4 (Figure R-14, indicated in blue rectangle). Furthermore, there were conserved residues (Figure R-14, highlighted in purple), which although they did not belong to a hinge region, they participated in interdomain interactions, formed in the vicinity of the hinge axes.

Sequence alignment revealed that although residues involved in hinges, were highly diverse, selected Ig-Ig transition motifs were conserved i.e. the NxxG motif, the presence of Leu residues at the end and of Pro residues in the beginning of almost each Ig domain. The aforementioned sequence features dictated the boundaries of Ig domains, which appeared to coincide for all the proteins examined, (including human DSCAM) independently of the residue type involved in the hinge bending mechanism. It was also concluded that the dynamic motion of every protein domain varies, based on the length of the hinge region. These findings suggested that the boundaries of DSCAM Ig1-Ig9 domains could be determined with potential hinges being intercalated between the Ig domains. Although the images produced from single-particle analysis suggested that hinge areas might be present among domains Ig5 to Ig8 (since domains Ig1-Ig5 formed a less flexible core), an accurate hinge location is difficult to be predicted considering only features based on the protein sequence.



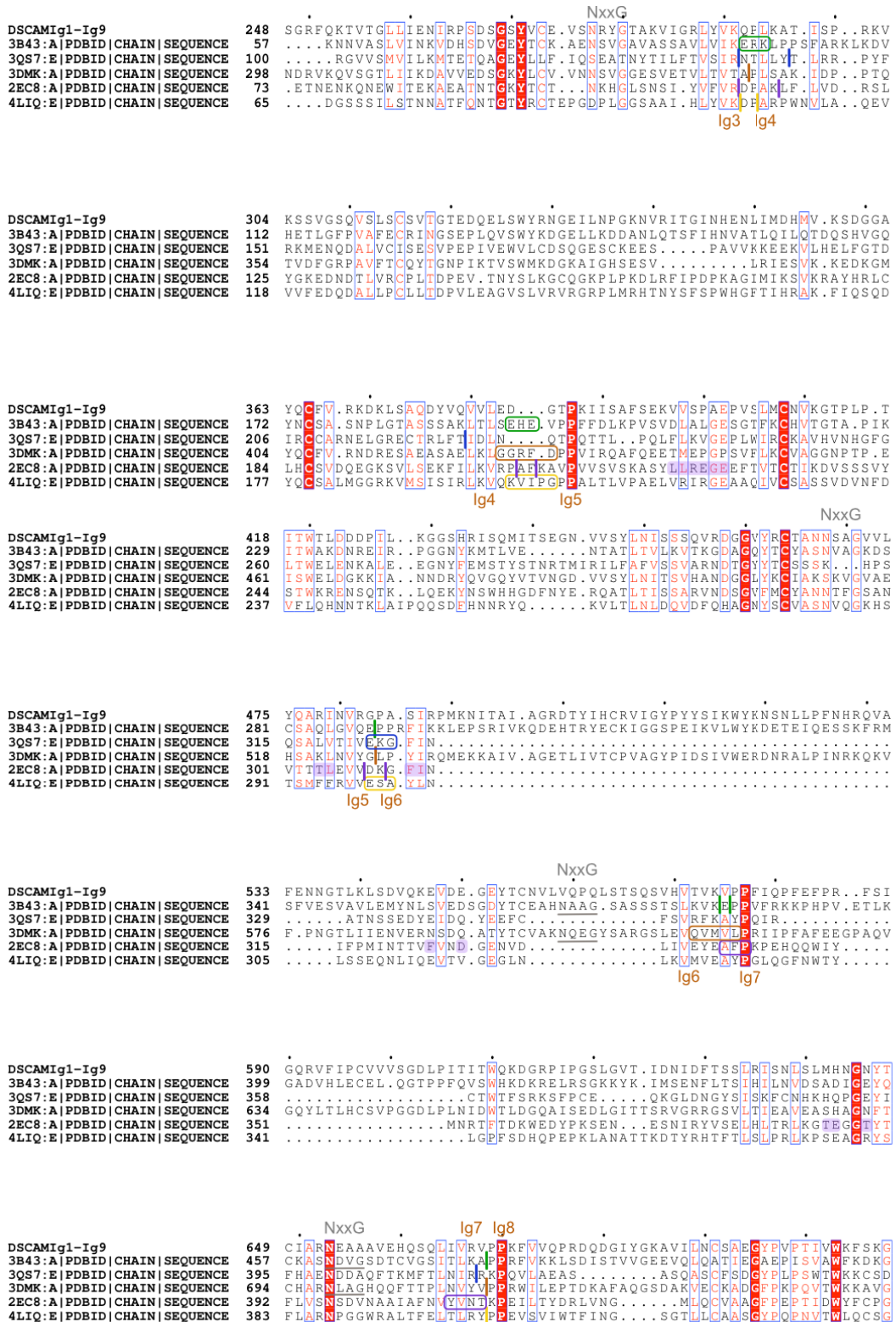


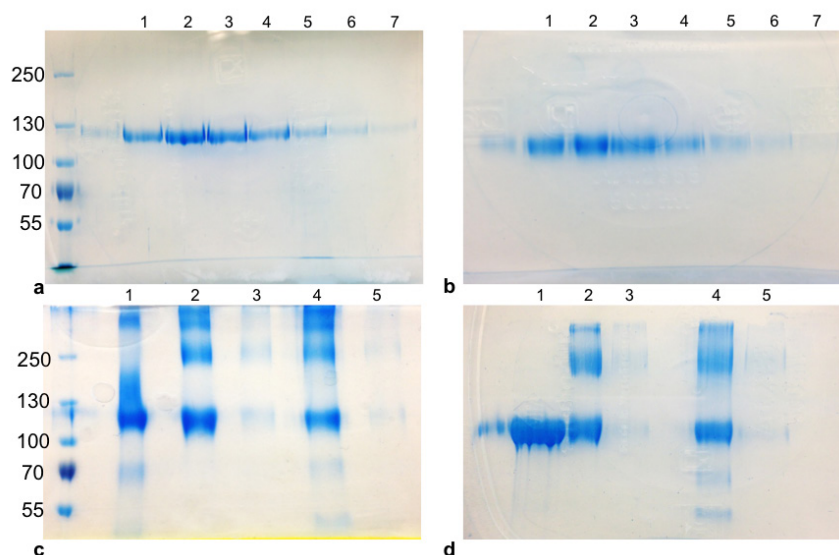
Figure R-14. Sequence alignment of DSCAM Ig1-Ig9 with protein receptors comprising Ig domains and hinge regions: Titin Ig65-Ig70 (25 % seq. homology, PDB entry 3B43) (green), Tyrosine kinase receptor 3-Flt3 Ig1-Ig4 (24 % seq. homology, PDB entry 3QS7) (blue), Dscam Ig1-Ig8 from *D. melanogaster* (34 % seq. homology, PDB entry 3DMK) (orange), Tyrosine kinase receptor-KIT Ig1-Ig5 (26 % seq. homology, PDB entry 2EC8) (purple), Colony-stimulating factor 1 (CSF-1) Ig1-Ig5 (22 % seq. homology, PDB entry 4LIQ) (yellow). Hinge regions for each protein are depicted in rectangles, the Ig domain boundaries with a vertical line, the conserved identical amino acids are white on a red background and similar

residues are red on a white background, following the color code given for the individual proteins. The highlighted amino acids (shown in purple) indicate the residues involved in Ig-Ig domain interactions. Multiple sequence alignment was performed using the ClustalW2 server (Sievers et al., 2011) and the output was processed and visualized using the ESPript 3.0 server (Robert and Gouet, 2014).

Single-particle analysis also revealed that the dimeric conformation was only rarely observed; therefore, DSCAM dimerization could not be elucidated (Figure R-12 marked in yellow). With the aim to increase the incidence of dimer observation, additional experiments were performed using a variety of salt concentrations and buffer solutions at different pH, but all conditions led to similar results with the ones previously described.

Further experiments then followed, giving emphasis on stabilizing the homodimer interactions through chemical crosslinking of DSCAM Ig1-Ig9 (WT and GnTI), prior to negative staining. To ensure a mild chemical reaction, an EM-grade glutaraldehyde solution (Table S2) was applied to DSCAM Ig1-Ig9 using the 'on-column' protocol (Shukla et al., 2014). A broad range of glutaraldehyde, salt concentrations and buffers were examined to establish the optimal conditions.

More specifically, crosslinking was performed with the WT protein using low glutaraldehyde concentrations (ranging from 0.1 to 0.8 % (v/v)) in PD-10 desalting columns, in the buffer solution after SEC (Table M-9). After elution, protein samples were evaluated with reduced and non-reduced SDS gels, which showed that the amount of glutaraldehyde used initially was not sufficient for protein crosslinking (Figure R-15a,b); hence an alternative and quick batch method was applied to crosslink the protein by incubating it for 5 min at 37 °C using 2.3 % (v/v) glutaraldehyde. Based on the MW of the dimer, the results demonstrated that some of the protein sample was crosslinked; however, non-specific crosslinked protein was also observed in higher molecular masses (Figure R-15c,d). Therefore, different glutaraldehyde concentrations were examined, using both the batch and the 'on-column' method to optimize the yield of the crosslinked protein. Despite the lower amount of crosslinked protein as compared to the one that was obtained by the batch method, the 'on-column' protocol was finally selected for protein crosslinking. This method would allow protein molecules subjected to non-specific crosslinking (with higher MW, prone to induce sample heterogeneity) to be discarded with mild crosslinking, using 1 and 2 % (v/v) glutaraldehyde.

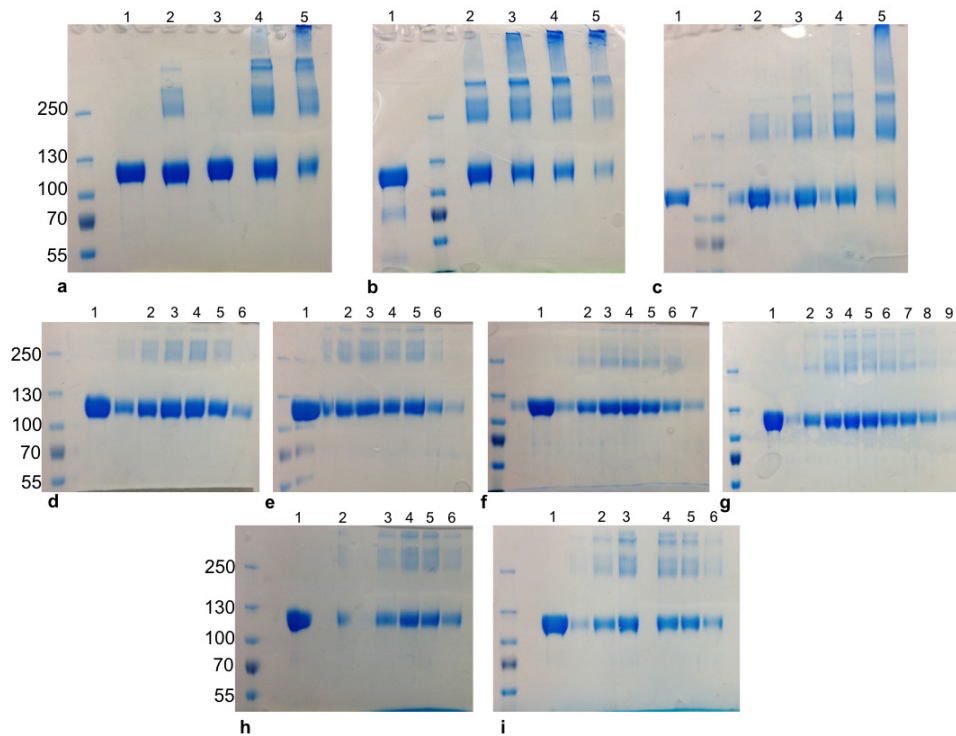


**Figure R-15.** DSCAM Ig1-Ig9 WT and GnTI<sup>+</sup> crosslinking. **a,b.** SDS-PAGE of DSCAM Ig1-Ig9 WT eluates (1 mg/ml) crosslinked in PD-10 columns, using 7 % reduced (**a**) and non-reduced gels (**b**). Buffer: 50mM MES pH 6.0, 150mM NaCl. Lane 1: control: non-crosslinked protein, lanes 2,3: eluates with 0.2 % (v/v) glutaraldehyde, lanes 4,5: eluates with 0.4 % (v/v) glutaraldehyde, lanes 6,7: eluates with 0.8 % (v/v) glutaraldehyde. **c,d.** SDS-PAGE of DSCAM Ig1-Ig9 WT and GnTI<sup>+</sup> (1 mg/ml) crosslinking in a batch method with 2.3 % (v/v) glutaraldehyde, using 7 % reduced (**c**) and non-reduced gels (**d**). Lane 1: WT control with non-crosslinked protein, lane 2: WT crosslinked in 50mM MES pH 6.0, 150mM NaCl, lane 3: WT crosslinked in PBS pH 7.4, lanes 4,5: GnTI<sup>+</sup> crosslinked in 50mM MES pH 6.0, 150mM NaCl and PBS pH 7.4, respectively.

The selection of the ‘on-column’ method proved to be correct since the chemical reaction was milder for DSCAM, as presented in [Figure R-16,a-c](#) for batch method and in [d-i](#) for the ‘on-column’ one.

In order to achieve separation between the crosslinked and non-crosslinked sample (resulted from the ‘on-column’ method), initially SEC was performed; however, the size differences of the two protein samples did not allow good resolution. Therefore, IEC was further tested with the anticipation that the chemical crosslinking would have altered the overall charge of the protein leading to sample separation. The elution profile showed that it was a mixture containing both protein populations. Nevertheless, the samples (for both WT and GnTI<sup>+</sup>) in the presence of NaCl (ranging from 0 to 1 M concentration) with 0.75 % uranyl formate, were further examined using EM negative staining and were imaged at RT. Protein destabilization and aggregation was observed when no salt was present in the buffer. The rest of the raw images indicated that the population of the crosslinked samples was more heterogeneous exhibiting high polydispersity, compared to the non-crosslinked ones.

Thus, it was concluded that single-particle analysis should only be conducted using the non-crosslinked population of DSCAM Ig1-Ig9.

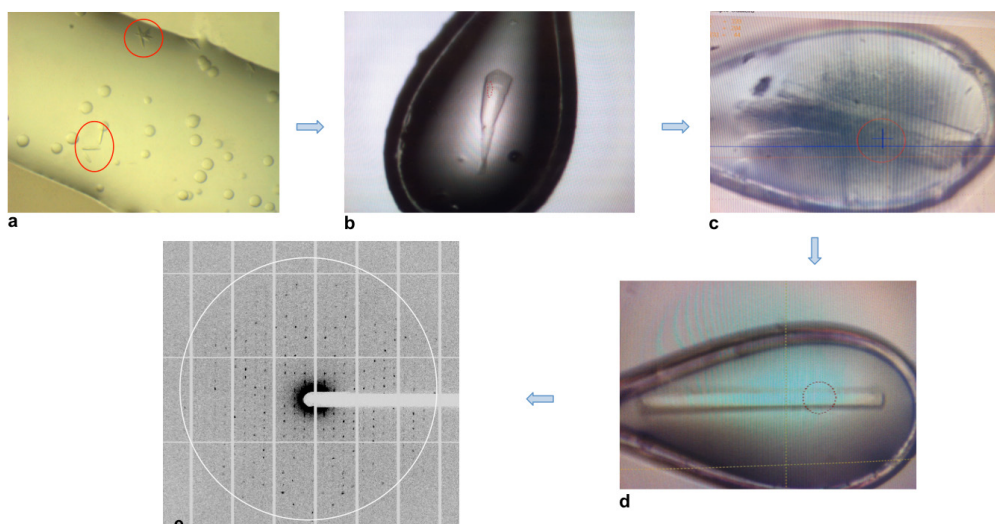


**Figure R-16.** DSCAM Ig1-Ig9 WT crosslinking. **a-c.** SDS-PAGE of DSCAM Ig1-Ig9 WT in a batch method using 1-5 % (v/v) glutaraldehyde in a range of buffers and salt concentrations. **a.** Buffer 20 mM MES pH 6.0, 50 mM NaCl with lanes 1,3: control with non-crosslinked protein sample and lanes 2,4,5: crosslinked protein with 1, 3, 5 % (v/v) glutaraldehyde, respectively. **b.** Buffer 20 mM Hepes pH 7.4, 50 mM NaCl with lane 1: control and lanes 2-5: crosslinked protein with 1, 2, 3, 5 % (v/v) glutaraldehyde, respectively. **c.** Buffer 50 mM MES pH 6.0, 150 mM NaCl with lane 1: control and lanes 2-5: crosslinked protein with 1, 2, 3, 5 % (v/v) glutaraldehyde, respectively. **d,e.** DSCAM Ig1-Ig9 WT eluates after crosslinking with PD-10 columns using 1 and 2 % (v/v) glutaraldehyde, respectively, in 20 mM MES pH 6.0, 50 mM NaCl with lane 1: control and lanes 2-6: crosslinked protein. **f,g.** DSCAM Ig1-Ig9 WT eluates after crosslinking with PD-10 columns using 1 and 2 % (v/v) glutaraldehyde, respectively, in 20 mM Hepes pH 7.4, 50 mM NaCl with lane 1: control and lanes 2-9: crosslinked protein. **h,i.** DSCAM Ig1-Ig9 WT eluates after crosslinking with PD-10 columns using 1 and 2 % (v/v) glutaraldehyde, respectively in 20 mM MES pH 6.0 and no additional salt with lane 1: control and lanes 2-6: crosslinked protein.

### 3.1.3.3 X-ray crystallography of DSCAM Ig1-Ig8 & Ig1-Ig9

Crystallization trials were set with DSCAM constructs expressed in WT and in GnTI<sup>-</sup> cells. Initially, the glycosylated protein (Ig1-Ig8 WT) was extensively tested for crystallization using a variety of conditions. However, the increased sample heterogeneity induced by the glycosylation sites probably prohibited the protein from crystallizing. To surpass that, the GnTI<sup>-</sup> form was used instead. Indeed, first crystals appeared in 400 nl drops (reservoir solution:protein volume ratio of 1:1), set up at the HTX facility (SPC, EMBL-Hamburg) under a variety of PEG conditions and buffers (PEG 4000, PEG 8000, PEG 20000, MES pH 6.0, Tris-HCl pH 8.0, Hepes pH 7.5) (Figure R-17a). Further optimization of the existing conditions was performed in larger volumes (1 - 2  $\mu$ l drops) using Linbro plates (vapor diffusion-sitting drop) by designing custom-made screens. The conditions explored aimed to examine the behavior of DSCAM in a range of pH, sample volume, concentration and mixing volume ratios with the reservoir solution in order to identify those conditions that would allow growing good quality diffracting crystals.

Crystals in different habits appeared in one week in the presence of PEG 4000 and MES pH 6.0 using 8 mg/ml of protein. The crystals were then exposed to X-rays but diffracted weakly to 18 Å resolution (Figure R-17b). Further optimization was performed using the additive screen (Hampton Research) where a range of ligands, organic solvents, salts and detergents was tested. The addition of 1,4-dioxane improved diffraction to 8 Å resolution (Figure R-17c) and emphasis was then given on changing the kinetics of vapor diffusion with the aid of oil. For that purpose, crystals were grown under oil in approximately two weeks and after exposure to the beam, resolution was slightly improved to 7 Å. A variety of cryo-protectants (glycerol, isopropanol, PEG 400, ethylene glycol) was also tested to examine their effect in crystal diffraction. A solution containing 20 % (v/v) ethylene glycol along with precipitant agents that the crystal was grown into, coupled with a soaking time of ~10 min, further improved the crystal diffraction and a data set was collected at 6.7 Å resolution (100 K) at PETRA III (Beamline P13, EMBL-Hamburg, DESY, Germany) (Figure R-17d,e). In parallel, additional methods to improve the diffraction quality of the crystal were examined, such as using the existing crystals as seeds (microseeding and seeding under oil); however, these seemed not to improve the diffraction power of the crystals. All the crystallization methods and optimization approaches used for obtaining crystals for DSCAM Ig1-Ig8 are summarized in Table R-6.



**Figure R-17.** Crystallization and optimization of DSCAM Ig1-Ig8 GnTI<sup>1</sup>. **a.** Thin needles developed in 15 - 30 % (w/v) PEG 8000/PEG 20000, 0.2 M MES pH 6.0/Tris-HCl pH 8.0 resulted in no diffraction. **b.** Single crystals with sharp, non-defined edges in 20 % (w/v) PEG 4000, 0.1 M MES pH 6.0 diffracted to 18 Å resolution. **c.** Single crystals with hollows at their growth ends in 15 % (w/v) PEG 4000, 0.1 M MES pH 6.0, 3 % (v/v) 1,4 dioxane diffracted to 8 Å resolution. **d.** Single 3D well-defined crystals, developed under oil (400 μl of silicon oil:paraffin oil (v/v) 40:60), in 15 % (w/v) PEG 4000, 0.1 M MES pH 6.0, 3 % (v/v) 1,4 dioxane using 20 % (v/v) ethylene glycol as cryo-protectant. **e.** Diffraction pattern of data collected from crystal shown in (d), at 6.7 Å resolution.

Diffraction images of DSCAM Ig1-Ig8 GnTI<sup>1</sup> collected at 6.7 Å resolution from a single crystal, were indexed and integrated with MOSFLM (Leslie et al., 2007) followed by POINTLESS (Evans, 2006) to determine the crystal lattice and symmetry. The integrated reflections were further scaled and merged using SCALA (Evans, 2006). Preliminary characterization demonstrated that the crystal belonged to C-centered orthorhombic lattice and C222<sub>1</sub> space group, with unit cell dimensions  $a = 118.6 \text{ \AA}$ ,  $b = 317.0 \text{ \AA}$ ,  $c = 283.5 \text{ \AA}$ ,  $\alpha = \beta = \gamma = 90.0^\circ$ . Data collection statistics are summarized in [Table R-2](#).

Table R-2. Data collection statistics of DSCAM Ig1-Ig8 GnTI<sup>-</sup>.

<b>Data collection and processing statistics</b>	
<b>Experiment</b>	DSCAM Ig1-Ig8 GnTI <sup>-</sup>
<b>X-ray source</b>	Beamline P13, PETRA III DESY, Germany
<b>Wavelength (Å)</b>	0.900
<b>No of images</b>	1800
<b>Oscillation angle (°)</b>	0.1
<b>Space group</b>	C222 <sub>1</sub>
<b>Unit cell dimensions (Å)</b>	a=118.6, b=317.0, c= 283.5, α=β=γ= 90.0°
<b>Resolution range (Å)</b>	25.0 - 6.7 (7.1 - 6.7) <sup>c</sup>
<b>No of observations</b>	62524 (9141)
<b>No of unique reflections</b>	9917 (1438)
<b>R<sub>m</sub><sup>a</sup></b>	0.090 (0.862)
<b>Completeness (%)</b>	99.6 (99.3)
<b>&lt;I/σ(I)&gt;<sup>b</sup></b>	11.6 (2.0)
<b>CC<sub>1/2</sub></b>	0.997 (0.790)
<b>Multiplicity</b>	6.4 (6.3)

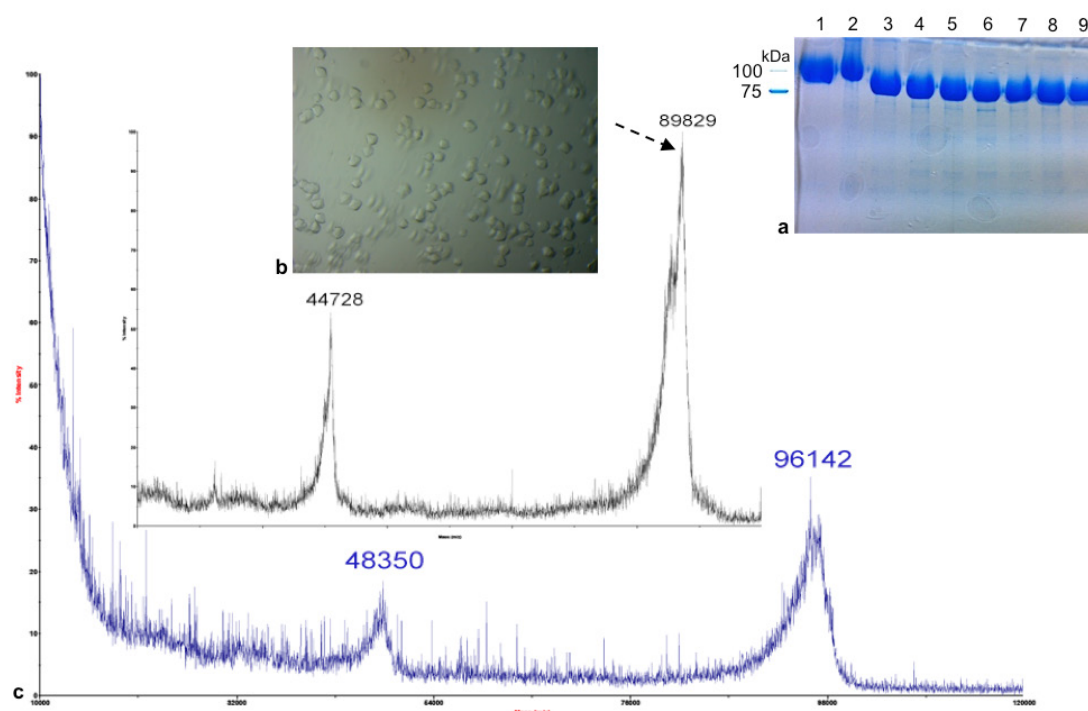
<sup>a</sup>  $R_m = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$ , where  $I_i(hkl)$  is the intensity of a reflection and  $\langle I(hkl) \rangle$  is the mean intensity of all  $i$  symmetry-related reflections.

<sup>b</sup>  $\sigma(I)$  is the standard deviation of  $I$ .

<sup>c</sup> Values in parentheses correspond to the outermost shell.

Efforts to increase the resolution of the diffraction data by post-crystallization treatments, focused on subsequent dehydration of the crystals, without achieving the desired results. Further studies were directed towards reducing the heterogeneity of the protein, prior to and during crystallization, by deglycosylation. To this end, the GnTI<sup>-</sup> protein was treated with EndoH so as to cleave asparagine-linked mannose oligosaccharides leaving one GlcNAc attached to Asn sites (§2.2.8.3e). DSCAM Ig1-Ig8 GnTI<sup>-</sup> was incubated with EndoH at RT, which resulted in reduction of its size. This was already observed after 15min of incubation time, as indicated by SDS-PAGE on samples taken at different time intervals (Figure R-18a). The previously identified crystallization conditions for the non-EndoH treated protein (Figure R-17d) were also tested with the deglycosylated sample (both by using the same conditions and by screening around them), without leading to crystal formation. Spare-matrix crystallization screening was also performed, using different conditions of PEG and pH, considering that deglycosylation would have altered the overall charge of the protein; however no crystals were formed. Alternatively, EndoH was incorporated in the crystallization drop (~3-5 %) repeating the sparse-matrix screening. The hypothesis was that using a lower enzyme concentration directly in the crystallization drop, would still lead to deglycosylated protein crystals, since the enzyme would interact longer with the protein. Successful crystallization of the deglycosylated

protein form, as indicated by MS, confirmed the aforementioned hypothesis. More specifically, cubic and tetragonal crystals were grown within three weeks in PEG 3350 and MES pH 6.5 (Figure R-18b). Verification of the MW of the crystallized protein treated with EndoH was conducted by MALDI-TOF analysis. The spectrum produced, revealed that some of the *N*-linked oligosaccharides have been truncated as indicated by the MW of the protein that was decreased by ~6 kDa compared to the MW of GnTI<sup>-</sup> protein, prior to enzyme's treatment (Figure R-18c). The deglycosylated crystals diffracted weakly to ~20 Å resolution.



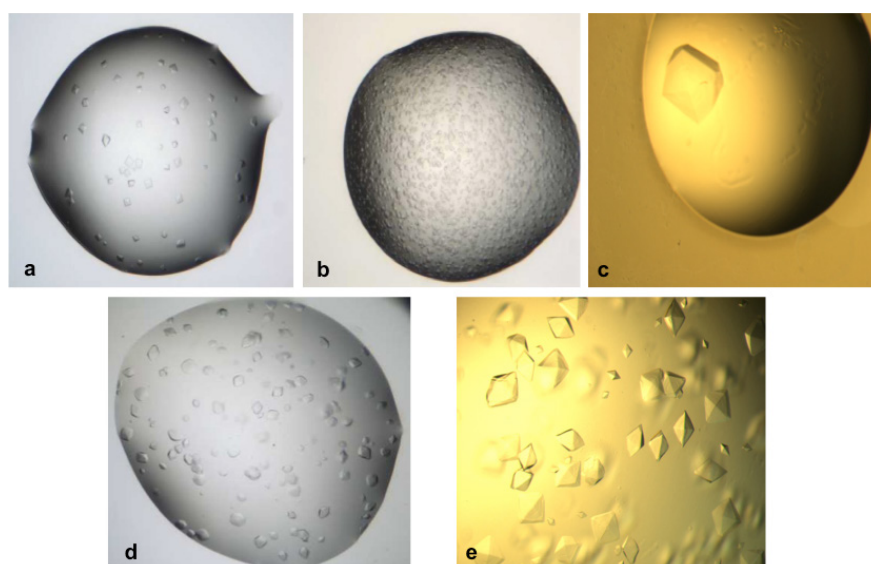
**Figure R-18.** EndoH treatment of DSCAM Ig1-Ig8 GnTI<sup>-</sup>. **a.** SDS-PAGE of deglycosylated DSCAM Ig1-Ig8 GnTI<sup>-</sup> when it was treated with EndoH at RT for 24h. Lane 1: control of DSCAM Ig1-Ig8 GnTI<sup>-</sup> without endoH, lane 2:  $t_0$ - immediately after EndoH addition, lane 3: 15 min incubation, lane 4: 30 min, lane 5: 60 min, lane 6: 2 h, lane 7: 4 h, lane 8: 8 h, lane 9: 24 h. **b.** Crystals of deglycosylated DSCAM Ig1-Ig8 GnTI<sup>-</sup> were grown within 3 weeks in 8% (w/v) PEG 3350, 50 mM MES pH 6.5 at 19 °C, only after the addition of EndoH in the crystallization drop (3 %). Crystals diffracted weakly to ~20 Å resolution. **c.** MALDI-TOF analysis of crystals grown in 3 % EndoH. The spectra in blue and black show the MW of DSCAM Ig1-Ig8 GnTI<sup>-</sup> prior to (~96 kDa) and after (~90 kDa) deglycosylation, respectively. The peaks corresponding to ~48 kDa and ~44 kDa, in each sample, are doubled charged states of the protein.

In order to decrease the flexibility of the amino acid chains exposed to the solvent, which could prevent the formation of well-ordered crystals, the effect of the protein's surface entropy reduction was also investigated. Reductive methylation of



lysine residues was performed in DSCAM Ig1-Ig8 GnTI<sup>+</sup> (~5 % Lys), where primary amines (Lys and N-terminal residues) were chemically modified to tertiary amines (Walter et al., 2006) (§2.2.8.3d). Although DSCAM was successfully methylated and crystals were grown, they diffracted to low-resolution (~7 Å).

DSCAM Ig1-Ig9, in contrast to Ig1-Ig8, was more prone to crystallize both in the WT and the GnTI<sup>+</sup> form, which was in agreement with what it was anticipated, based on biophysical results (monodisperse population compared to DSCAM Ig1-Ig8). Crystallization trials of DSCAM Ig1-Ig9 were set up at the HTX facility in small-scale (300 nl drops) using a range of volume ratios between the reservoir solution and the protein sample. The optimum starting protein concentration for crystallization was estimated with pre-crystallization tests by Hampton Research. Initial crystal hits for both WT and GnTI<sup>+</sup> were observed under various conditions of PEG and other precipitant agents using commercial screens from Molecular Dimensions and Qiagen (Figure R-19a,b,d). These conditions were further optimized in large-scale (2 µl drops, 1:1 volume ratio in 500 µl reservoir solution) and crystals were grown within one week at 19 °C, as shown in Figure R-19c,e, using the vapor diffusion method (hanging drop). The crystals were tested at ID29 beamline (ESRF, Grenoble, France) and they exhibited weak diffraction to ~9 Å resolution.



**Figure R-19.** Crystallization of DSCAM Ig1-Ig9 WT and GnTI<sup>+</sup>. **a.** Cubic and rhombic crystals of WT set up at the HTX facility, which were grown within 8 days, in the presence of 0.2 M potassium thiocyanate, 0.1 M sodium cacodylate pH 6.5, 8 % (w/v) PEG 20000, 8 % (w/v) PEG 550 MME using 7.5 mg/ml protein. **b.** Small round crystals of WT with non-defined edges set up at the HTX facility in the presence of 1 M Lithium chloride, 0.1 M MES pH 6.0, 10 % (w/v) PEG 6000 were grown in 2 days using 7.5 mg/ml protein. **c.** Optimization of crystallization conditions (crystals shown in image b) resulted in a single large tetragonal

crystal, grown in 0.8 M Lithium chloride, 0.1 M MES pH 6.0, 7 % (w/v) PEG 6000 in 2  $\mu$ l drops (reservoir solution:protein 1:2) using 5 mg/ml of DSCAM Ig1-Ig9 WT. **d.** Round crystals with non-defined edges of GnTI<sup>-</sup> (12 mg/ml) were grown an the HTX facility in 4 days, in 0.2 M potassium thiocyanate, 0.2 M potassium bromide, 0.1 M sodium cacodylate pH 6.5, 5 % (w/v) PEG 8000, 3 % (w/v) poly- $\gamma$ -glutamic acid of low MW polymer (PGA-LM). **e.** Optimization of crystals shown in image d: Custom-made plates in 2  $\mu$ l drops (reservoir solution:protein 1:1). The crystals had clear edges and they were grown in 0.2 M potassium thiocyanate, 0.1 M potassium bromide, 0.1 M sodium cacodylate pH 6.5, 9 % (w/v) PEG 8000, 2 % (w/v) PGA-LM using 10 mg/ml of DSCAM Ig1-Ig9 GnTI<sup>-</sup>.

Efforts to obtain crystals diffracting to a resolution that would allow at least the detection of the binding domains upon DSCAM homodimerization in vertebrates, proved to be rather challenging. Despite the different approaches employed for controlling the rate of crystal growth and the different optimization methods used, the diffraction quality and the mosaic spread of the crystals ( $\sim 1.0^\circ$ ) could not be further improved.

### 3.1.4 Binding affinity studies and *in vitro* reconstitution of DSCAM Ig1-Ig9 WT:netrin<sub>VIV</sub>

DSCAM Ig1-Ig9 binding affinity to netrin<sub>VIV</sub> and their stoichiometry were examined using ITC. Netrin<sub>VIV</sub> exhibited a high propensity to aggregate in a variety of buffer solutions, salt concentrations and pH; therefore, obtaining an optimal stabilization buffer for both proteins proved to be a challenge. The buffer solutions that were selected, were those that showed reduced precipitation of netrin<sub>VIV</sub>, i.e. MES pH 6.0, NaCl, DTT, CaCl<sub>2</sub>, (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> or PBS pH 8.0 (§2.2.7.4). Nevertheless, the results obtained from the assays in the presence of these solutions showed that netrin<sub>VIV</sub> precipitated in the ITC cell during the experiments.

Experiments were also performed aiming reconstitution of DSCAM Ig1-Ig9:netrin<sub>VIV</sub> complex *in vitro*, with co-purification of the proteins using size exclusion chromatography. Each component was purified separately and incubation of DSCAM Ig1-Ig9 with netrin<sub>VIV</sub> followed, in 1:1 and 1:2 molar ratios, at 4 °C for 30-60 min. The protein mixture was further purified using SEC in the presence of both buffers and the purification profile resulted in two single peaks that corresponded to DSCAM Ig1-Ig9 and netrin<sub>VIV</sub>, indicating that the complex could not be formed (Figure R-20).

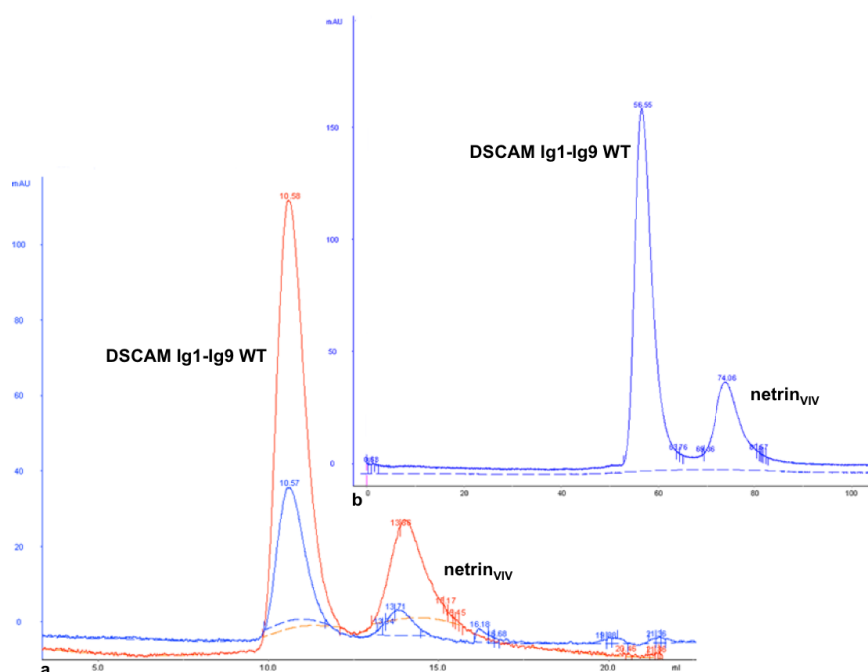
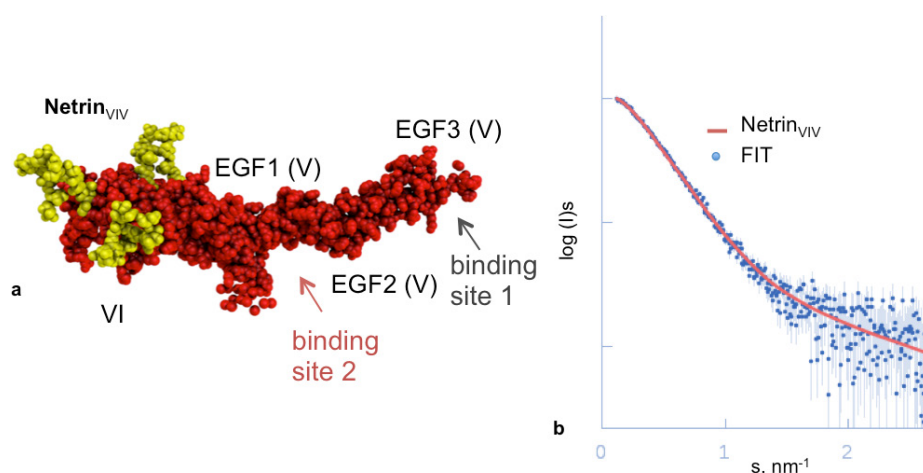


Figure R-20. Co-purification profile of DSCAM Ig1-Ig9 WT with netrin<sub>VIV</sub> after SEC in **a.** 1:1 molar ratio (DSCAM Ig1-Ig9:netrin<sub>VIV</sub>) in 50 mM MES pH 6.0, 150 mM NaCl, 1mM DTT, 2mM CaCl<sub>2</sub>, 1mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> shown in red and in PBS pH 8.0 shown in blue. **b.** 1:2 molar ratio (DSCAM Ig1-Ig9:netrin<sub>VIV</sub>) in PBS pH 8.0. The peak that elutes first corresponds to DSCAM Ig1-Ig9 and the second peak corresponds to netrin<sub>VIV</sub>.

### 3.1.5 Structural characterization of DSCAM Ig1-Ig9 WT:netrin<sub>VIV</sub>

#### 3.1.5.1 Small-angle X-ray scattering of DSCAM Ig1-Ig9 WT:netrin<sub>VIV</sub>

Structural characterization and shape determination of DSCAM Ig1-Ig9 WT in association with netrin<sub>VIV</sub> was conducted with SAXS. Since prior information regarding the stoichiometry of netrin and DSCAM could not be extracted from binding assays or co-purification of the complex (where the concentration range was much lower than the one used for SAXS), DSCAM was incubated with netrin<sub>VIV</sub> prior to measurements in 1:1 molar ratio at RT for 15 min in a concentration range of 0.6-4.0 mg/ml. All measurements were performed in collaboration with Dr. Haydyn Mertens (Svergun Group) at beamline P12 (PETRA III, EMBL-Hamburg, DESY, Germany). The data obtained from the measurements were compared to the SAXS data acquired from the individual components, where netrin<sub>VIV</sub> (Finci et al., 2014) (Figure R-21) was characterized as a monomer and DSCAM Ig1-Ig9 as a dimer in solution (Figure R-10).



**Figure R-21.** Netrin<sub>VIV</sub> analysis derived from SAXS experimental data. **a.** Rigid body modeling of netrin<sub>VIV</sub> (shown in red) using atomic coordinates from its crystal structure (PDB entry 4URT). The glycosylation sites are depicted in yellow. **b.** The calculated scattering curve (shown in red) fits well to the experimental data (shown in blue) and is plotted as the logarithm of scattered intensity against the momentum transfer ( $s$ ), with the quality of fit,  $\chi^2$ , being at 1.1.

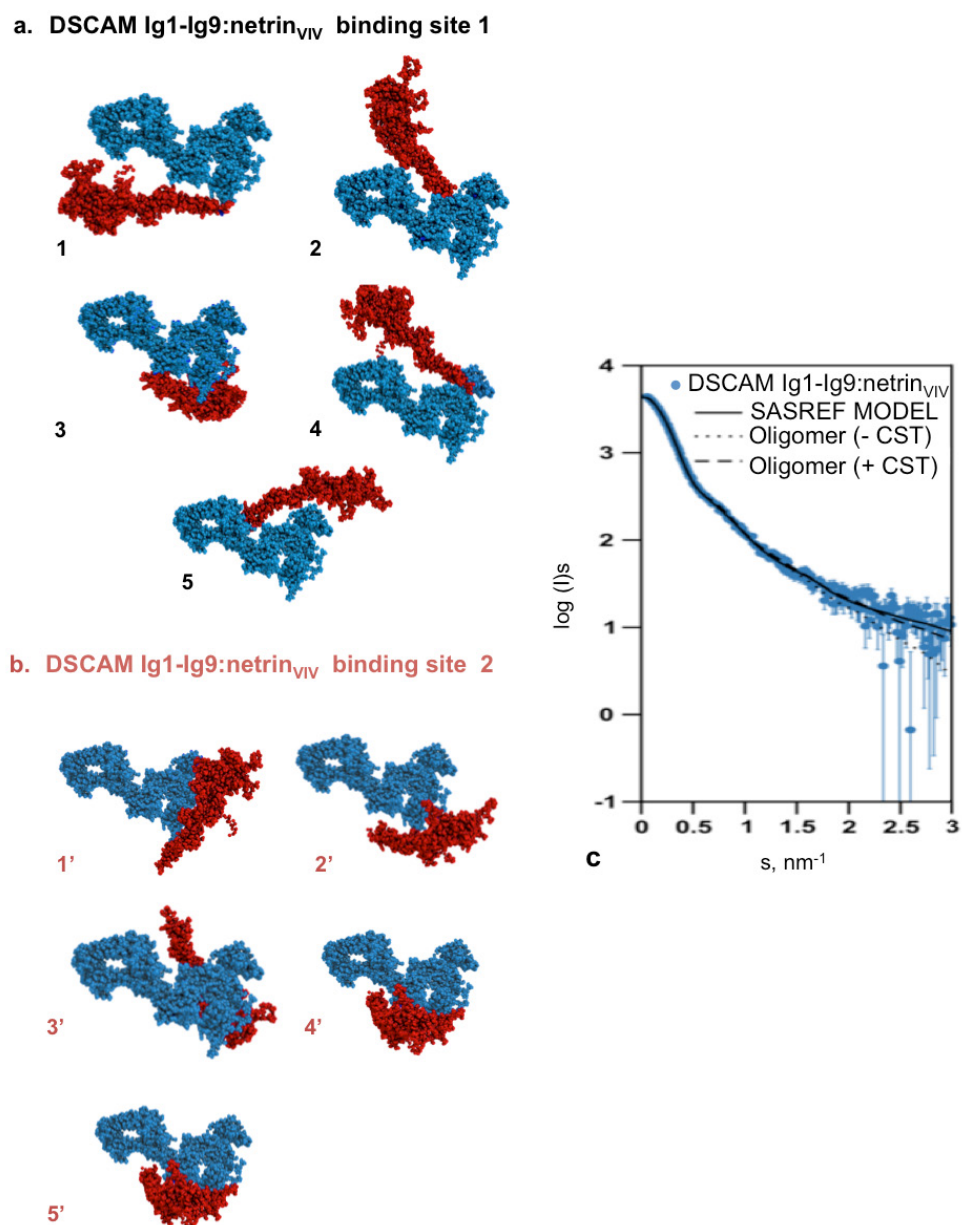
SAXS data were analyzed using the existing crystal structure of netrin<sub>VIV</sub> along with rigid body models of DSCAM Ig1-Ig9, (generated in the absence of netrin<sub>VIV</sub> §3.1.3.1), in an equilibrium analysis using the program OLIGOMER (Konarev et al., 2003). Analysis was performed in order to determine the fractions of the different

oligomeric states in the samples. The volume fractions of the population corresponded to a mixture of 70 % DSCAM Ig1-Ig9 dimer, 13 % DSCAM Ig1-Ig9 monomer and 17 % free netrin<sub>VIV</sub> with an overall data fit being at 1.4 ( $\chi^2 \sim 1.4$ ), suggesting that the two proteins did not form a complex.

However, the SAXS parameters extracted directly from the experimental data, e.g. the decrease in the  $D_{\max}$  and the MW calculations of DSCAM Ig1-Ig9:netrin<sub>VIV</sub>, which showed a lower than the expected value, compared with the ones for the DSCAM Ig1-Ig9 dimer when alone in solution, suggested that these proteins could be alternatively engaged in complex formation. The MW estimation for DSCAM Ig1-Ig9:netrin<sub>VIV</sub> could be attributed to one monomer of DSCAM and one monomer of netrin<sub>VIV</sub> (Table R-3). Based on these observations, rigid body models were constructed with SASREF (Petoukhov et al., 2005) using DSCAM Ig1-Ig9 monomer (extracted from CORAL dimeric model, as previously described) and netrin<sub>VIV</sub> monomer (extracted from the crystal structure with DCC56, PDB entry 4URT) (Finci et al., 2014). The latter has shown that netrin<sub>VIV</sub> occupies two distinct binding sites of DCC56 with binding site 1 being at the tip of netrin<sub>VIV</sub> involving EGF3 domain and binding site 2, involving EGF1 and EGF2 domains. DCC56 and DSCAM Ig1-Ig9 share certain amino acids, which are present in both binding sites of netrin<sub>VIV</sub>. Previous work has shown that DSCAM Ig1-Ig9 binds to netrin-1 (Ly et al., 2008); thus, the hypothesis that DSCAM could potentially interact with the same sites as DCC was examined. Although, it was proposed that binding site 2 could accommodate other receptors as well (Finci et al., 2014), both binding sites of netrin<sub>VIV</sub> were examined with rigid body modeling.

A radius of 10 Å was selected for probing DSCAM Ig1-Ig9 WT surface with netrin<sub>VIV</sub> and an ensemble of models was generated for both binding sides that fit equally the experimental data with  $\chi^2$  value being 1.2 (Figure R-22). Based on the quality of the fits, the models for both sites were complied with the experimental data; however, in order to define the site that was more favorable energetically, a statistical analysis was performed based on the abundance of different amino acid pair interactions in known structures in the Protein Data Bank (PDB). Distance-dependent pair-potentials have been generated based on the frequency of observed amino acids interactions by using Boltzmann formulation. These statistical interaction potentials enable fast energy calculations and ranking of binding sites (Svergun Group, manuscript in preparation). The scoring of the two binding sites of DSCAM Ig1-Ig9:netrin<sub>VIV</sub> in the heterocomplex is shown in Table R-4. The orientation that

seemed most favorable energetically for binding site 1 was for models 1 and 3 and for binding site 2 the best scoring orientations were generated for models 2', 4' and 5'. However, it is not clear whether DSCAM Ig1-Ig9 interactions with netrin<sub>VIV</sub> are limited to binding sites 1 and 2 due to the resolution limit imposed by SAXS data collection.



**Figure R-22.** a,b. DSCAM Ig1-Ig9 WT:netrin<sub>VIV</sub> rigid body models in complex formation (blue for DSCAM Ig1-Ig9 and red for netrin<sub>VIV</sub>) reconstructed using SASREF (Petoukhov et al., 2005). Both netrin<sub>VIV</sub> binding sites for DCC, may accommodate DSCAM Ig1-Ig9 equally using a contact constrain of 10 Å. c. The calculated scattering curve (shown in black) fits well the experimental scattering data (shown in blue) with  $\chi^2$  values of 1.2. Data are plotted as the logarithm of scattered intensity against the momentum transfer (s).

**Table R-3.** Table of selected statistics for DSCAM Ig1-Ig9, netrin<sub>VIV</sub> and DSCAM Ig1-Ig9:netrin<sub>VIV</sub> derived from SAXS data collection. The complete table of statistics can be found at the Appendix Section.

Protein	$R_g$ Guinier (nm)	$D_{max}$ (Å)	Volume (nm <sup>3</sup> )	MW <sup>1</sup> (kDa)	MW <sup>2</sup> (kDa)	MW <sup>3</sup> (Da)	Oligomeric state
DSCAM Ig1-Ig9	6.4±0.2	22.3±0.5	726±70	226±20	207±20	96890	Dimer
Netrin <sub>VIV</sub>	3.9±0.1	13.5±0.5	100±14	48±5	52±5	49289	Monomer
DSCAM Ig1-Ig9:netrin <sub>VIV</sub>	6.3±0.2	20.5±0.5	474±45	178±15	163±15	146179	-

MW<sup>1</sup>: calculation based on the absolute scattering intensity  $I(0)$ .

MW<sup>2</sup>: calculation based on BSA (used as a standard).

MW<sup>3</sup>: calculation based on sequence (without glycosylation).

**Table R-4.** Ranking of DSCAM Ig1-Ig9:netrin<sub>VIV</sub> models for binding site 1 and binding site 2.

Model	Z-score binding site1	Z-score binding site 2
1	2.76	
2	1.85	
3	2.42	
4	2.03	
5	1.91	
1'		2.06
2'		2.57
3'		2.48
4'		2.56
5'		2.56

### 3.1.5.2 Electron microscopy of DSCAM Ig1-Ig9 WT:netrin<sub>VIV</sub>

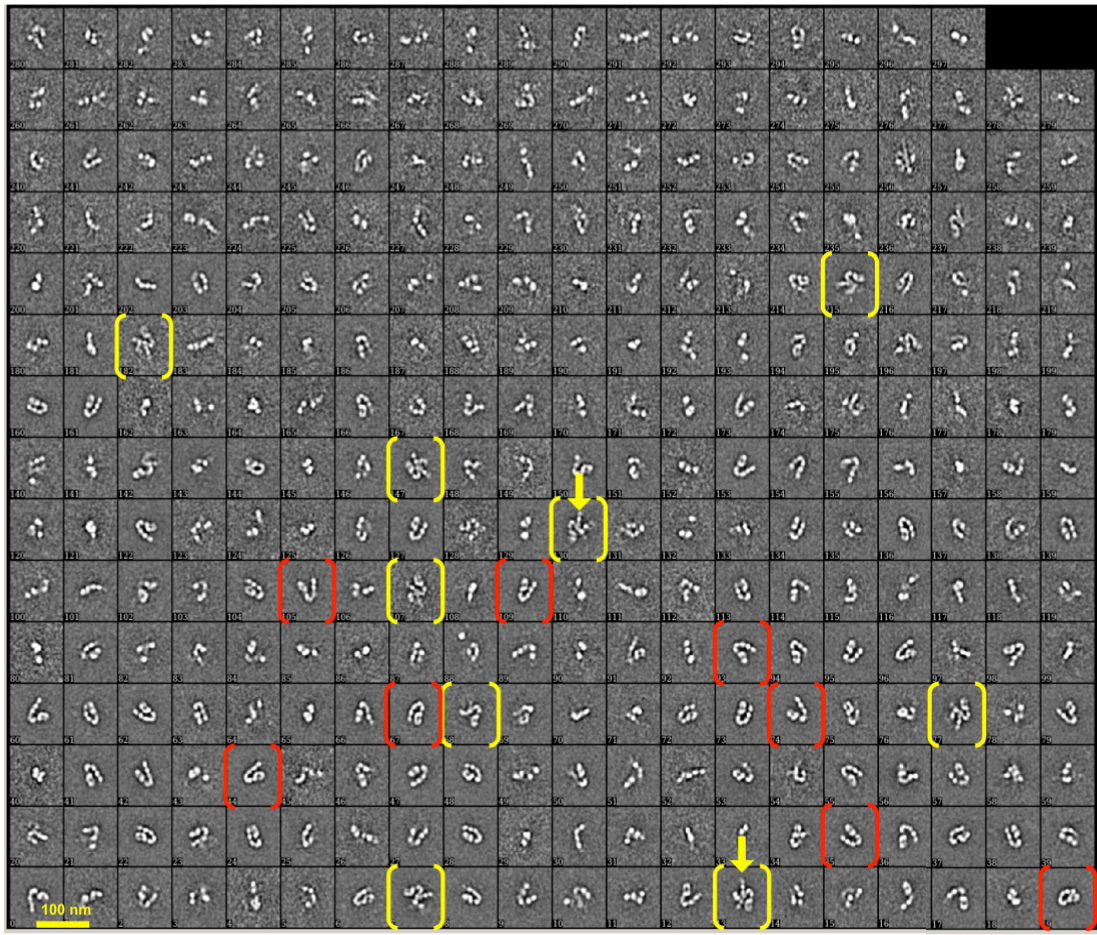
With the aim to investigate the interactions formed between DSCAM Ig1-Ig9 and netrin<sub>VIV</sub>, electron microscopy with negative staining experiments were performed by Skiniotis Group (Life Sciences Institute, University of Michigan). Prior to staining, DSCAM was mixed with netrin<sub>VIV</sub> in 1:1 molar ratio and incubated for ~15 min at RT.

Analysis of the raw images showed an increase in the number of larger-in-size complexes, compared to the ones obtained from DSCAM images examined in the absence of netrin<sub>VIV</sub>. Reference-free alignment and classification followed in order to obtain two-dimensional averages of the complex. A total of 12,236 particle projections of DSCAM with netrin<sub>VIV</sub> were subjected to ISAC to distinguish the population between the complex and the individual components, resulting in 298 classes with 20 rounds of classification. These results further supported that DSCAM was bound to a second molecule (Figure R-23).

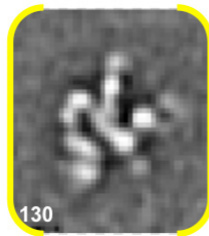
More specifically, single-particle analysis of DSCAM Ig1-Ig9 WT with netrin<sub>VIV</sub> in 1:1 molar ratio, showed that only a few projections presented DSCAM (consisting of domains Ig1-Ig9) interacting with a second molecule (Figure R-23a marked in yellow), while in most of the projections DSCAM Ig1-Ig9 monomer was primarily visualized having the same domain arrangement to the one adopted without netrin<sub>VIV</sub> (§3.1.3.2) (Figure R-23a marked in red). Netrin<sub>VIV</sub> with a MW of 56 kDa, consists of 420 amino acids and comprises a laminin domain VI and three EGF-like hands (Figure R-23d,e). Although it was anticipated that the second molecule would be netrin<sub>VIV</sub>, the EM findings showed a larger-in-size partner molecule interacting with DSCAM. The possibility that the second molecule could be netrin<sub>VIV</sub> cannot be excluded; however, based on the shape and size of the domains of the second partner, which were similar to DSCAM Ig folding, it was suggested that the partner molecule was only partially observed and could be attributed to DSCAM. A close-up view of selected images is presented in Figure R-23b,c.

In order to confirm the identity of the second partner and resolve the discrepancy, efforts focused towards increasing the number of images that the complex was visualized. Thus, further experiments were performed by increasing the concentration of netrin during incubation with DSCAM (DSCAM Ig1-Ig9:netrin<sub>VIV</sub> molar ratio in 1:2). The results obtained from negative staining, showed similar 2D projections of the molecules with the ones recorded where 1:1 molar ratio was used, without providing additional information on the second binding molecule. Selected projections are shown in Figure R-24.

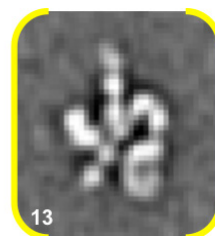




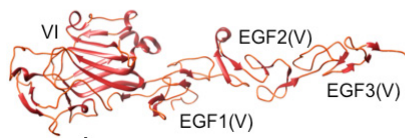
a



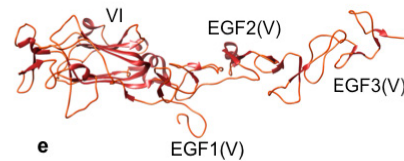
b



c



d



e

**Figure R-23. a.** Single-particle analysis of DSCAM Ig1-Ig9 WT with netrin<sub>IV</sub> in 1:1 molar ratio, in 50 mM MES pH 6.0, 150 mM NaCl, 1mM DTT, 2mM CaCl<sub>2</sub>. Reference-free alignment and classification of particle projections was performed with ISAC. 12,236 particle projections of DSCAM with netrin<sub>IV</sub> were subjected to ISAC resulting in 298 classes from 20 rounds of classification. A DSCAM monomer (Ig1-Ig9) can be primarily visualized (shown in red). A few averages show DSCAM forming larger complexes (marked in yellow) where DSCAM Ig1-Ig9 is fully visualized with the second partner being partially observed. **b,c.** Close-up view of selected projections of DSCAM Ig1-Ig9 dimer formation. Front (**d**) and side view (**e**) of netrin<sub>IV</sub> crystal structure obtained at 3.1 Å (Finci et al., 2014) comprising the laminin domain VI and three EGF-like domains.

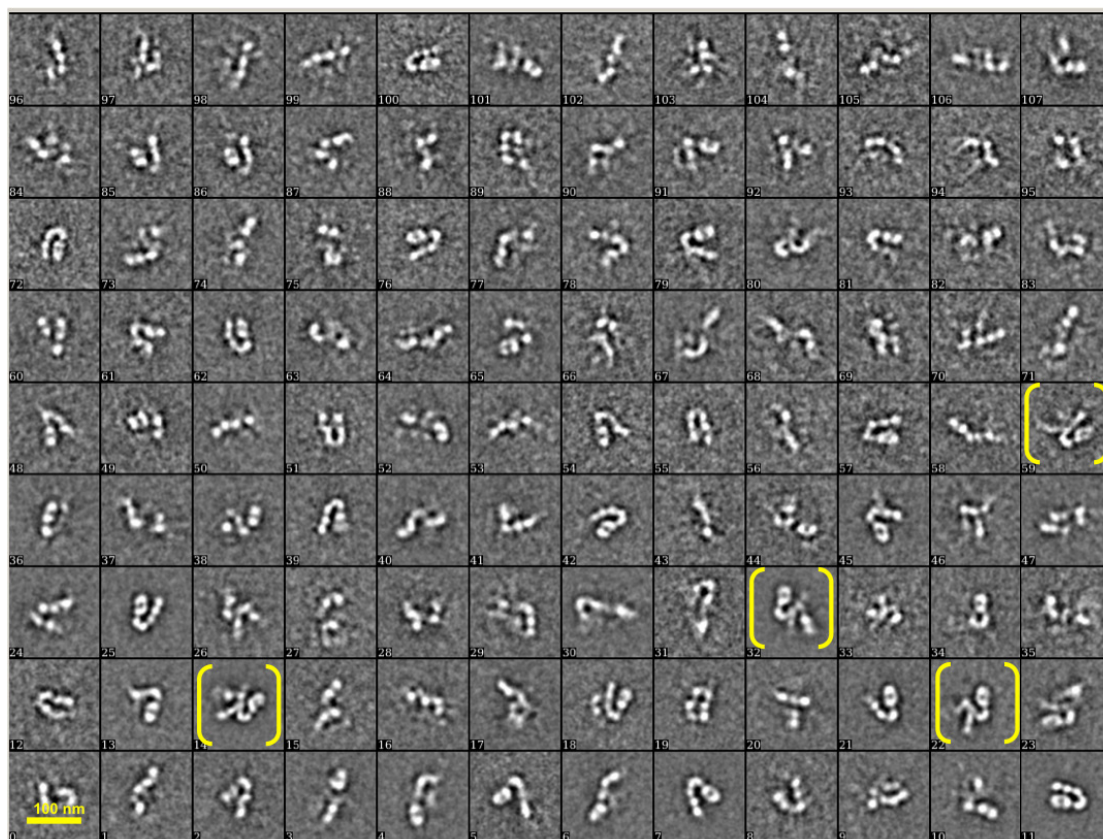


Figure R-24. Single-particle analysis of DSCAM Ig1-Ig9 WT with netrin<sub>VIV</sub> in 1:2 molar ratio. 7919 particle projections of DSCAM with netrin<sub>VIV</sub> were subjected to ISAC resulting in 108 classes from 20 rounds of classification. Most of the projections resulted in monomeric DSCAM whereas the rest presented DSCAM forming dimers with the second molecule being partially observed.

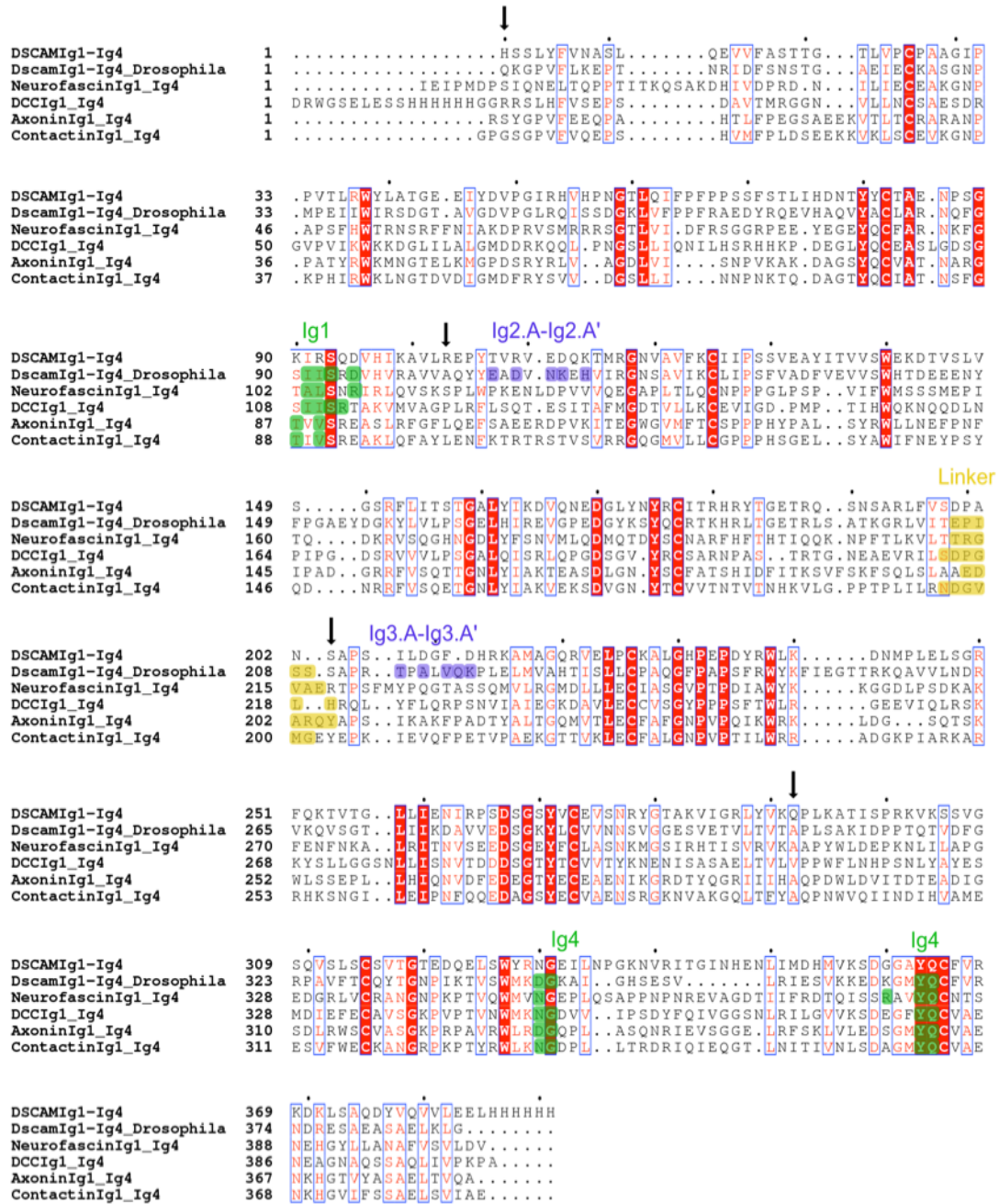
The ambiguity introduced by the second molecule observed interacting with DSCAM could be possibly resolved by crystallographic studies. However, in the frame of the present thesis, further experiments of DSCAM Ig1-Ig9:netrin<sub>VIV</sub>, were not conducted.

## **DSCAM Ig1-Ig4**

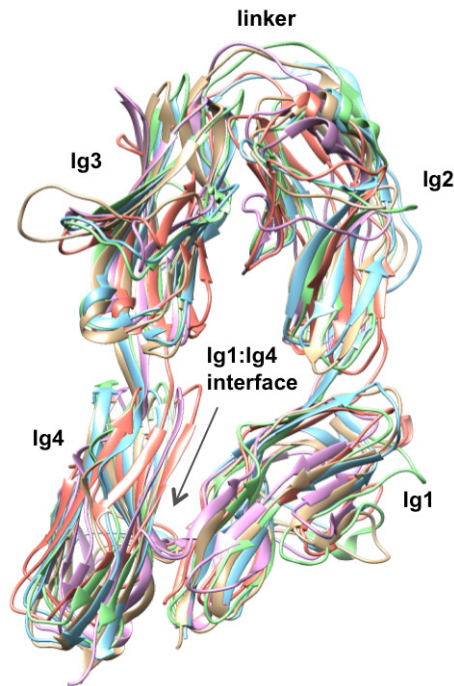
### **3.1.6 Construct design rationale of DSCAM Ig1-Ig4**

The results obtained from EM experiments for the longer DSCAM construct (Ig1-Ig9), suggested that a rigid horseshoe-like configuration was adopted for the first four *N*-terminal Ig domains. This horseshoe-shape resembled the arrangement of Dscam in *Drosophila*, corresponded to the same Ig domains (Meijers et al., 2007) and it was also a common feature observed in other proteins involved in neuron recognition (Figure R-26). To this direction, DSCAM human sequence (Ig1-Ig4) was aligned with sequence homologous proteins with known crystal structures i.e. *N*-terminal DCC Ig1-Ig4 (Chen et al., 2013), *N*-terminal fragment of axonin-1/TAG-1 (Freigang et al., 2000), contactin4 Ig1-Ig4 (Bouyain and Watkins, 2010) and neurofascin homophilic adhesion molecule (Liu et al., 2011) in order to identify common sequence features that could explain the horseshoe configuration (Figure R-25).

According to previously determined structures, domains Ig1-Ig2 and Ig3-Ig4 form tight connections without the intervener of a linker; however there is a loop region consisting of five or six residues between domains Ig2-Ig3 that facilitates the horseshoe configuration (Figure R-25, highlighted in yellow). Although, the residues comprising the linker are hypermutable, there is a number of conserved amino acids that contribute to the U-shape (horseshoe) Ig1-Ig4 arrangement. This U-shape is further stabilized through conserved residues involved in intramolecular interactions creating an interface for Ig1:Ig4. More specifically, in the case of Dscam, residues Asp344 and Gly345 from domain Ig4, form hydrogen bonds with Ser93 from domain Ig1. In addition, Gln369 from domain Ig4 interacts with Ile91 from domain Ig1, creating a bridge between domains Ig1 and Ig4 that secures the structure. In close proximity to the aforementioned residues, certain conserved residues contribute as well to Ig1:Ig4 intraface, involved in hydrophobic interactions (Figure R-25, highlighted in green). The aforesaid amino acids are also conserved in human DSCAM forming a stable core structure with domains Ig1 and Ig4 to be engaged in a more rigid substructure. These findings were in agreement with the conclusions made from EM studies regarding the shape of this smaller construct. Therefore, the hypothesis that this structure would be more prone to crystallize compared to the longer constructs was tested and a smaller DSCAM Ig1-Ig4 construct was additionally designed.



**Figure R-25.** Sequence alignment of human DSCAM Ig1-Ig4 with Dscam in *D. melanogaster*. 32 % sequence identity (over 124 out of 391 amino acids), with DCC: 26 % sequence identity (over 89 out of 343 amino acids), with axonin4: 26 % sequence identity (over 90 out of 352 amino acids), with contactin4: 26 % sequence identity (over 89 out of 349 amino acids) and neurofascin: 23 % sequence identity (over 86 out of 378 amino acids). The arrows indicate the boundaries of the Ig domains based on Dscam sequence (Ig1: 1-102, Ig2: 103-204, Ig3: 210-305, Ig4: 306-388) and the linker connecting domains Ig2-Ig3 is highlighted in yellow (a.a.: 205-209). The conserved identical residues for the six proteins are white on a red background and similar residues are red on a white background. The conserved amino acids involved in the Ig1:Ig4 intraface are highlighted in green. The amino acids involved in the dimer interface between Ig2.A-Ig2.A' and Ig3.A-Ig3.A' in Dscam in *Drosophila*, are highlighted in purple. Multiple sequence alignment was performed using the ClustalW2 server (Sievers et al., 2011) and the output was processed and visualized using the ESPript 3.0 server (Robert and Gouet, 2014).



**Figure R-26.** Superposition of Dscam Ig1-Ig4 structure, in monomer, (shown in orange) (PDB entry 2V5M, chain A) with structural homologous proteins adopting a horseshoe-like configuration. In purple: DCC FN1-FN4 (PDB entry 3LAF, chain A), in beige: axonin-1/TAG-1 (PDB entry 1CS6, chain A), in light blue: contactin4 (PDB entry 3KLD, chain A), in green: neurofascin (PDB entry 3P3Y, chain A). Superposition was conducted with Lsqkab (CCP4 suite) (Kabsch, 1976) over  $C_{\alpha}$  atoms as matching residues and the molecular graphics representation was performed with UCSF CHIMERA (Pettersen et al., 2004).

Recent findings emerging from studies on Dscam Ig1-Ig4 crystal structures in *D. melanogaster* using eight isoforms, in a resolution range from 1.9 Å to 4.0 Å, confirmed the aforementioned findings and suggested that all isoforms adopt a horseshoe module with the same conserved amino acids contributing to the Ig1:Ig4 intraface, stabilizing the whole Dscam arrangement. In addition, the amino acids involved in the dimer interface formed between Ig2-Ig2 and Ig3-Ig3 are all conserved among the different isoforms examined, including the isoform representing the structure with PDB ID 2v5m (Li et al., 2016); however these residues are not conserved in human DSCAM (Figure R-26, highlighted in purple).

### 3.1.7 Purification & biophysical characterization of DSCAM Ig1-Ig4

DSCAM Ig1-Ig4 was cloned into the pXLG vector equipped with a His<sub>6</sub>-tag at the C-terminus, as described in §2.2.3 (primers listed in Table S5). Cloning was followed by transformation into DH5a *E. coli* cells and after sequence validation the protein was expressed using HEK 293S and GnTI<sup>-</sup> cells in one Liter of medium, secreted

using its native secretion signal. The protein expression was evaluated by WB (Figure R-27a).

DSCAM Ig1-Ig4 was purified with IMAC followed by SEC (Figure R-27b) resulting in ~20 mg from one Liter of cell culture for WT (and in ~10-15 mg for the GnT<sup>I</sup> protein). According to elution profile, using a Hiload Superdex 75 prep grade column, DSCAM Ig1-Ig4 was eluted in a single peak. It was estimated that the protein was in monomeric state under the conditions used for purification, in contrast to the longer constructs (DSCAM Ig1-Ig8 & Ig1-Ig9).

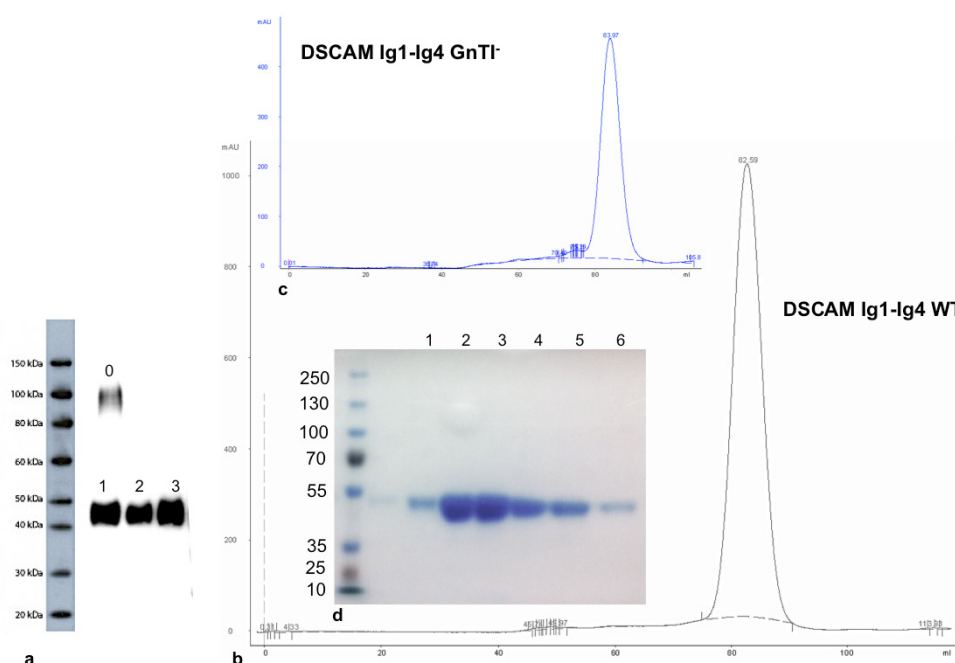
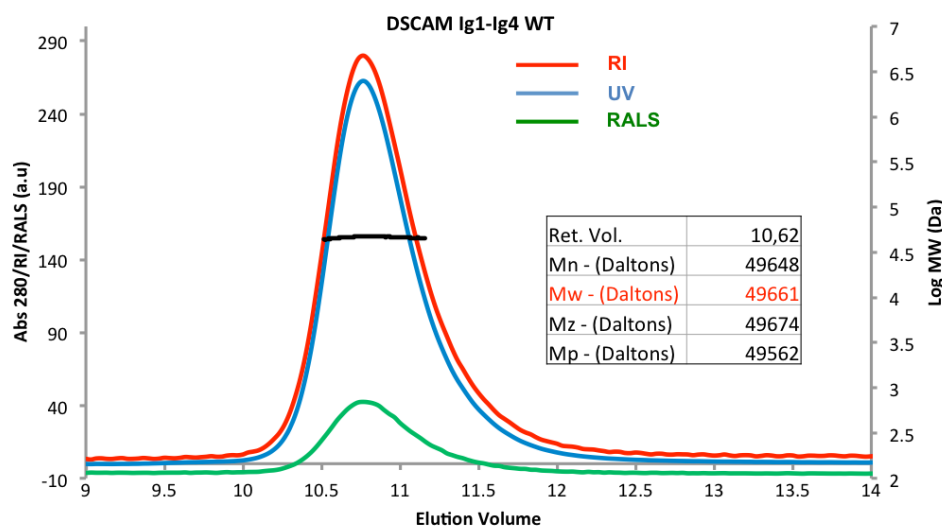


Figure R-27. DSCAM Ig1-Ig4 WT and GnT<sup>I</sup> expression and purification profile: **a**. Evaluation of expression by WB, lane 0: control of DSCAM Ig1-Ig8 expression, lanes 1,2: expression of DSCAM Ig1-Ig4 WT in different media preparations, lane 3: expression of DSCAM Ig1-Ig4 GnT<sup>I</sup>. **b,c**. Purification profile after SEC with the protein being eluted in a single peak for DSCAM Ig1-Ig4 WT and GnT<sup>I</sup>, respectively. **d**. Evaluation of DSCAM Ig1-Ig4 WT purity with SDS-PAGE, lanes 1-6: correspond to the eluates after SEC.

To further confirm the oligomeric state of DSCAM Ig1-Ig4, analysis of the WT form was performed with HPLC-RALS/RI/UV using the Malvern's OmniSEC GPC/SEC multidetector system (§2.2.7.3). The protein was loaded on a Superdex 10/300 GL column and it was eluted with a stable flow rate of 0.4 ml/min. The MW distribution of the peak corresponded to monomeric state and agreed with the aforementioned results from the purification profile (Figure R-28). The MW was  $49\pm 3$  kDa and was determined by UV measurements derived from RALS intensities and also by

independent comparison with BSA, which was used as standard protein. The MW determination was consistent with the results from the MS analysis, where it was found that the MW of the protein was ~48 kDa for the WT and ~47 kDa for the GnTI<sup>-</sup>. In parallel, the same experiment was performed for the GnTI<sup>-</sup> protein, which was also found to be a monomer in solution.



**Figure R-28.** SEC coupled with SLS detector of DSCAM Ig1-Ig4 WT. The oligomeric state analysis after SEC, demonstrated that it was a monomer in solution with a molecular mass of  $49 \pm 3$  kDa.

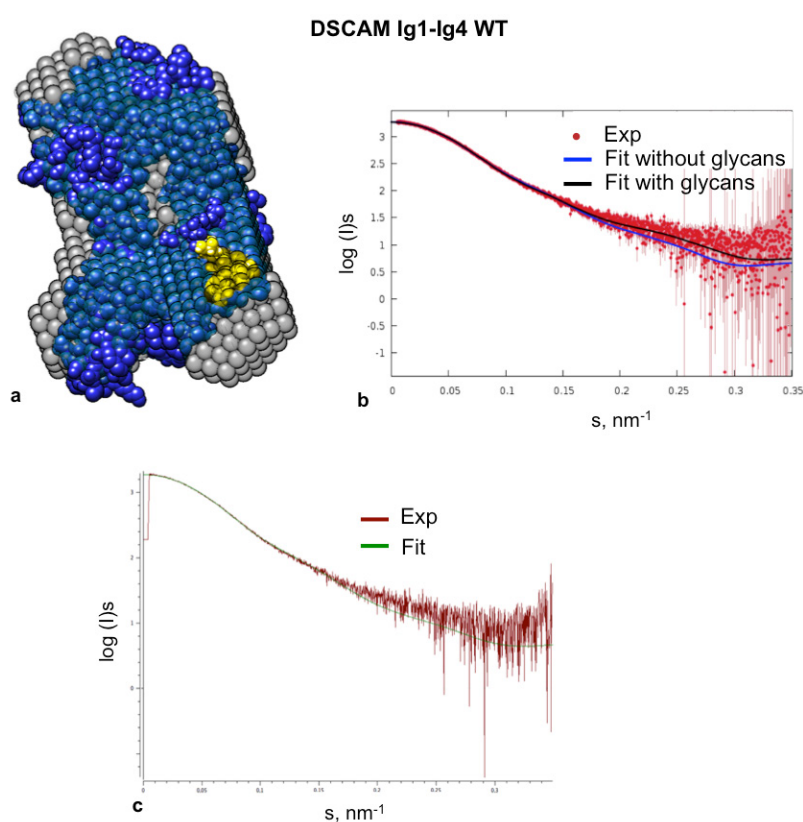
### 3.1.8 Structural characterization of DSCAM Ig1-Ig4

#### 3.1.8.1 Small-angle X-ray scattering of DSCAM Ig1-Ig4

SAXS measurements were performed for DSCAM Ig1-Ig4 WT in order to obtain information for the protein structural features and its overall shape in solution. A series of protein concentrations were examined (0.5 - 10.1 mg/ml) to identify changes in oligomeric state as concentration was increased. In addition, a number of salt concentrations were tested (0 mM - 1 M NaCl) to explore the effect of salt in DSCAM oligomeric state. The analysis demonstrated that the state of the protein in solution was monomeric, independent of protein concentration or of the presence of the salt in buffer solution.

Models were generated using *ab initio* and rigid body modeling methods. More specifically, the ensemble of *ab initio* models indicated that the four Ig domains adopted a monomeric conformation with domain arrangement resembling the Dscam structure in *Drosophila* (Meijers et al., 2007) (Figure R-26). The results were

complied with rigid body models calculated independently where the Ig domains adopted a horseshoe configuration and superposition of the *ab initio* and rigid body models resulted in normalized spatial discrepancy value (NSD) of 1.0 suggesting a good agreement between the two methods (Figure R-29a). The calculated scattering data from rigid body modeling fit to the experimental scattering curve with calculated discrepancy value ( $\chi^2$ ) of 1.1 (Figure R-29b). SAXS experimental data for human DSCAM Ig1-Ig4 showed good correlation in comparison with the crystallographic structure from *Drosophila* (Dscam Ig1-Ig4, PDB entry 2V5M), supporting the working hypothesis for this construct, as previously described in §3.1.6 (Figure R-29c).



**Figure R-29.** SAXS analysis and model generation of DSCAM Ig1-Ig4 WT. **a.** Superposition of the *ab initio* model (transparent beads) with the model obtained from rigid body modeling (shown in dark blue) demonstrated good agreement between the two methods (NSD= 1.0). The theoretical N-glycosylation sites on DSCAM Ig1-Ig4 model (in yellow) were built using the GlyProt server (Bohne-Lang and Von der Lieth, 2005) based on DSCAM amino acid sequence. **b.** Scattering curve of the calculated data from rigid body modeling with and without glycans (shown in black and blue, respectively) that fit the experimental scattering curve (shown in red). The quality measure of the fit of the simulated scattering curve to the experimental data ( $\chi^2$ ) was 1.3 for the model without glycans and 1.1 for the glycosylated model, indicating that the fit was better when glycans were introduced in the model. **c.** The theoretical scattering curve (shown in green) calculated from the crystal structure of Dscam Ig1-Ig4 (*D. melanogaster*) fits well to the SAXS experimental data (shown in red) with  $\chi^2$  of 1.1. Data are plotted as the logarithm of scattered intensity against the momentum transfer ( $s$ ).



**Table R-5.** Table of selected statistics for DSCAM Ig1-Ig4 derived from SAXS data collection. The complete table of statistics can be found at the Appendix Section.

<b>Protein</b>	<b><math>R_g</math> Guinier (nm)</b>	<b><math>D_{max}</math> (Å)</b>	<b>Volume (nm<sup>3</sup>)</b>	<b>MW<sup>1</sup> (kDa)</b>	<b>MW<sup>2</sup> (kDa)</b>	<b>MW<sup>3</sup> (Da)</b>	<b>Oligomeric state</b>
DSCAM Ig1-Ig4	2.9±0.1	10±0.1	71±5	53±5	51±5	43754	Monomer

MW<sup>1</sup>: calculation based on the absolute scattering intensity  $I(0)$ .

MW<sup>2</sup>: calculation based on BSA (used as a standard).

MW<sup>3</sup>: calculation based on sequence (without glycosylation).

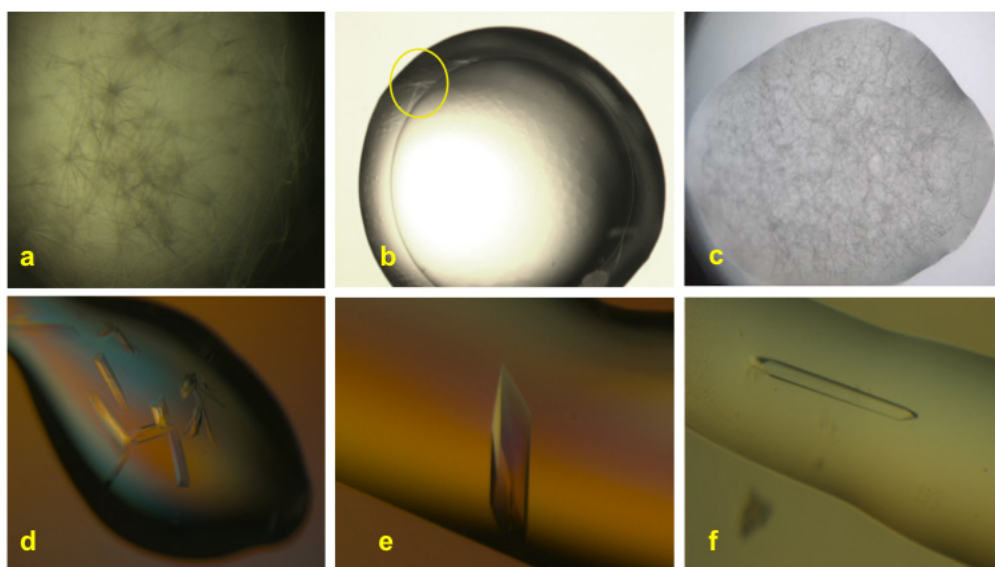
### 3.1.8.2 X-ray crystallography of DSCAM Ig1-Ig4

DSCAM crystallization was performed for both WT and GnTI<sup>-</sup> protein using a variety of commercially available screens from Qiagen. Drops (volume in the range of 150-300 nl) were automatically set up utilizing the robotic systems at the HTX facility (SPC, EMBL-Hamburg) and a variety of mixing volume ratios (reservoir solution to protein ratio 1:1, 1:2 and 2:1) were examined. Initial hits included clusters of microcrystals (DSCAM Ig1-Ig4 GnTI<sup>-</sup>, 10 mg/ml) grown in PEG 3350 and tri-sodium citrate in 2:1 protein:reservoir solution drops, using the sitting drop vapor diffusion method (Figure R-30a). Based on these hits, custom-made screens were designed using larger volumes (1-2  $\mu$ l) to optimize the existing conditions (using both the hanging and the sitting drop method); however, the crystal size remained small. Additional optimization trials were conducted employing microseeding. A series of dilutions of microcrystals (seeds) that were granulated from larger crystals by vortexing, were prepared and the seeds were introduced into a fresh protein-precipitant droplet of similar conditions to those that crystals were originally grown. Microseeding was coupled with adjustments in protein concentration, precipitant solution, pH and storage temperature. Nevertheless, only microcrystals were grown (Figure R-30b). In addition, the additive screen from Hampton was also used, but no further improvement of the crystal size was achieved.

After 60 days of incubation time at 19 °C, small thin needles of DSCAM Ig1-Ig4 appeared for the WT form, in 96-well plates that were set up at the HTX facility in the presence of PEG 4000, sodium cacodylate pH 6.5 and PGA-LM (Figure R-30c). These conditions were further optimized using larger drop volumes (1-2  $\mu$ l drops in 500  $\mu$ l precipitant solution) in 24-well Linbro plates with the vapor diffusion method (hanging and sitting drop). Large crystals, in different habits, appeared in three days in the presence of PEG 3350, sodium cacodylate pH 6.5 and PGA-LM using 10 mg/ml of DSCAM Ig1-Ig4 WT, with various types of visible defects at their edges,

which diffracted to 10 Å resolution (Figure R-30d). Prior to data collection, crystals were soaked in a solution consisting of the reservoir ingredients with a small increase in PEG 3350 concentration (3-5 % (w/v)) and with the addition of 20 % (v/v) glycerol or 25 % (v/v) ethylene glycol for cryoprotection. Data collection that was also performed at RT showed that the cryoprotectant did not affect the crystal diffraction.

Additional improvement of the existing crystals was performed by employing the under oil-crystal growth using the same conditions, with incorporation of the cryoprotectant (20 % glycerol or ethylene glycol) in the reservoir solution (Figure R-30e,f). The protein:reservoir solution volume ratio was 1:1 and the crystals were grown within seven to ten days. This method improved the crystal size and quality and a complete data set was collected to 2.7 Å resolution. The crystallization methods and optimization approaches used for obtaining crystals for all DSCAM constructs are summarized in Table R-6.



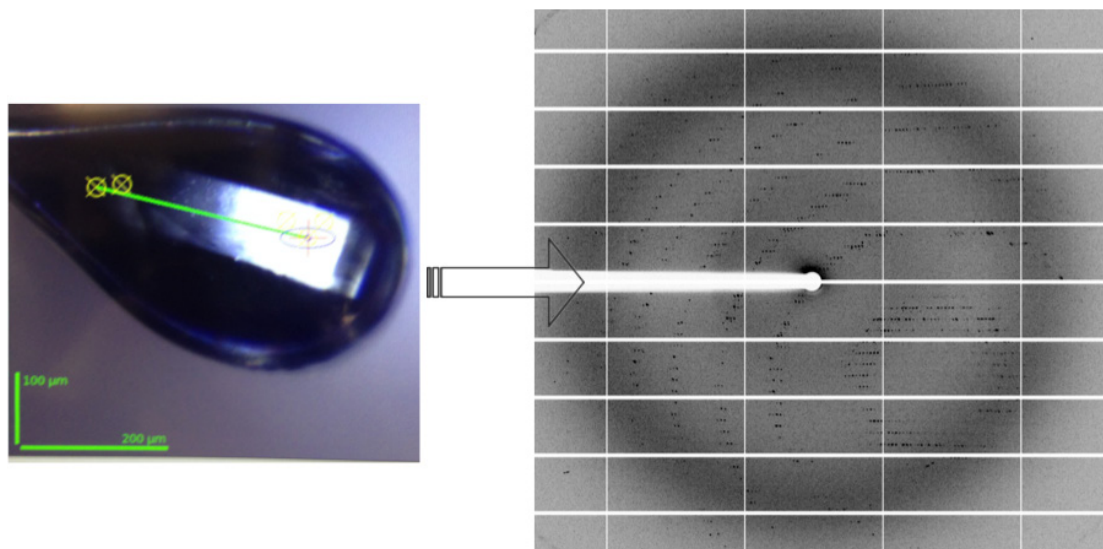
**Figure R-30.** Crystallization and optimization of DSCAM Ig1-Ig4 GnTI<sup>-</sup> (a,b) and WT (c-f). **a.** Thin microcrystals (DSCAM Ig1-Ig4 GnTI<sup>-</sup>) in inseparable clusters in 20 % (w/v) PEG 3350 and 0.2 M tri-sodium citrate in 2:1 protein:reservoir solution drops set up by robotic systems. **b.** Thin needles after microseeding in 15 % (w/v) PEG 3350 and 0.1 M sodium citrate. **c.** Cluster of thin needles (DSCAM Ig1-Ig4 WT) set up at the HTX facility, appeared in 60 days in 15 % (w/v) PEG 4000, 0.1 M sodium cacodylate pH 6.5 and 5 % (v/v) PGA-LM. **d.** Optimization of conditions from image c, led to larger crystals in different morphologies with non-defined edges (13-20 % (w/v) PEG 3350, 0.1 M sodium cacodylate pH 6.5 and 2-5 % (v/v) PGA-LM). **e,f.** Crystals developed under oil (500 µl silicon oil:paraffin oil (v/v) 50:50) in 15 % (w/v) PEG 3350, 0.1 M sodium cacodylate pH 6.5 and 2 % (v/v) PGA-LM where 20 % (v/v) glycerol or ethylene glycol was incorporated in the reservoir solution during crystal growth, leading to single 3D crystals without any visible cracks or defects.

**Table R-6.** Summary of the crystallization methods and optimization approaches employed for all DSCAM constructs.

	DSCAM Ig1-Ig8 WT	DSCAM Ig1-Ig8 GnTI <sup>+</sup>	DSCAM Ig1-Ig9 WT	DSCAM Ig1-Ig9 GnTI <sup>+</sup>	DSCAM Ig1-Ig4 WT	DSCAM Ig1-Ig4 GnTI <sup>+</sup>
Sparse Matrix screening (vapor diffusion)	✓	✓	✓	✓	✓	✓
Vapor diffusion Hanging drop/ Sitting drop (custom-made plates)	✓/-	✓/✓	✓/-	✓/-	✓/✓	✓/✓
Microbatch	-	✓	-	-	-	-
Hanging drop under oil	-	✓	-	-	✓	-
Methylation	-	✓	-	-	-	-
EndoH treatment	-	✓	-	-	-	-
Temperature variation	-	✓	-	-	-	✓
Additives	-	✓	-	-	-	✓
Seeding & matrix seeding	✓	✓	-	-	-	✓
Seeding under oil	-	✓	-	-	-	-
Dehydration	-	✓	✓	✓	-	-
Crystals/ Resolution	-	✓/6.7 Å	✓/9 Å	✓/9 Å	✓/2.7 Å	✓/.*

\* No diffraction

Data collection of DSCAM Ig1-Ig4 WT (Figure R-30e,f) was performed at ESRF (Beamline ID29, Grenoble, France) and a complete data set was collected from a single crystal at 2.7 Å resolution using a helical data collection (Figure R-31). The reflections were indexed and integrated with XDS (Kabsch, 2010), followed by POINTLESS (Evans, 2006) to determine the crystal lattice and symmetry. The integrated reflections were then scaled and merged using SCALA (Evans, 2006). The crystal belonged to C-centered orthorhombic lattice, space group C222<sub>1</sub> with unit cell dimensions a= 63.3 Å, b= 139.7 Å, c= 226.7 Å,  $\alpha=\beta=\gamma= 90.0^\circ$ . Data collection and statistics are summarized in Table R-7.



**Figure R-31.** DSCAM Ig1-Ig4 WT helical data collection at ID29 (ESRF, Grenoble, France) of a single crystal grown under oil in 15 % (w/v) PEG 3350, 0.1 M sodium cacodylate pH 6.5 and 2% (v/v) PGA-LM with cryoprotectant incorporation in the reservoir solution (20 % (v/v) glycerol) during crystal growth. A full data set was collected at 2.70 Å and the crystal belonged to space group C222<sub>1</sub>.

**Table R-7.** Data collection statistics of DSCAM Ig1-Ig4 WT on C222<sub>1</sub>.

<b>Data collection and processing statistics</b>	
<b>Experiment</b>	DSCAM Ig1-Ig4 WT
<b>X-ray source</b>	Beamline ID29 ESRF, Grenoble, France
<b>Wavelength (Å)</b>	0.976
<b>No of images</b>	1800
<b>Oscillation angle (°)</b>	0.1
<b>Space group</b>	C222 <sub>1</sub>
<b>Unit cell dimensions (Å)</b>	a=63.3, b=139.7, c= 226.7 $\alpha=\beta=\gamma= 90.0^\circ$
<b>Resolution range (Å)</b>	25.0 - 2.7 (2.9 - 2.7) <sup>c</sup>
<b>No of observations</b>	158258 (21955)
<b>No of unique reflections</b>	24382 (3437)
<b>R<sub>m</sub><sup>a</sup></b>	0.075 (0.378)
<b>Completeness (%)</b>	96.7 (94.2)
<b>&lt;I/σ(I)&gt;<sup>b</sup></b>	17.0 (4.3)
<b>CC<sub>1/2</sub></b>	0.998 (0.948)
<b>Multiplicity</b>	6.5 (6.4)
<b>B-factor (Å<sup>2</sup>) (Wilson Plot)</b>	55.8

<sup>a</sup>  $R_m = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$ , where  $I_i(hkl)$  is the intensity of a reflection and  $\langle I(hkl) \rangle$  is the mean intensity of all  $i$  symmetry-related reflections.

<sup>b</sup>  $\sigma(I)$  is the standard deviation of  $I$ .

<sup>c</sup> Values in parentheses correspond to the outermost shell.

A complete data set was also collected at 3.1 Å resolution, from crystals grown under the same conditions, with ethylene glycol instead of glycerol, being incorporated in the precipitant solution. These crystals belonged to Primitive

orthorhombic lattice, space group  $P2_12_12_1$  with unit cell dimensions  $a = 66.7 \text{ \AA}$ ,  $b = 147.1 \text{ \AA}$ ,  $c = 230.1 \text{ \AA}$ ,  $\alpha = \beta = \gamma = 90.0^\circ$ . Data collection and statistics are summarized in Table R-8.

Table R-8. Data collection statistics of DSCAM Ig1-Ig4 WT on  $P2_12_12_1$ .

<b>Data collection and processing statistics</b>	
<b>Experiment</b>	DSCAM Ig1-Ig4 WT
<b>X-ray source</b>	Beamline ID29 ESRF, Grenoble, France
<b>Wavelength (Å)</b>	0.976
<b>No of images</b>	1800
<b>Oscillation angle (°)</b>	0.1
<b>Space group</b>	$P2_12_12_1$
<b>Unit cell dimensions (Å)</b>	$a = 66.7$ , $b = 147.1$ , $c = 230.1$ $\alpha = \beta = \gamma = 90.0^\circ$
<b>Resolution range (Å)</b>	25.0 - 3.1 (3.3 - 3.1) <sup>c</sup>
<b>No of observations</b>	271649 (35706)
<b>No of unique reflections</b>	42127 (5911)
<b><math>R_m</math><sup>a</sup></b>	0.095 (0.736)
<b>Completeness (%)</b>	99.4 (97.2)
<b><math>\langle I/\sigma(I) \rangle</math><sup>b</sup></b>	12.1 (2.5)
<b>CC<sub>1/2</sub></b>	0.997 (0.927)
<b>Multiplicity</b>	6.4 (6.0)
<b>B-factor (Å<sup>2</sup>) (Wilson Plot)</b>	75.0

<sup>a</sup>  $R_m = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$ , where  $I_i(hkl)$  is the intensity of a reflection and  $\langle I(hkl) \rangle$  is the mean intensity of all  $i$  symmetry-related reflections.

<sup>b</sup>  $\sigma(I)$  is the standard deviation of  $I$ .

<sup>c</sup> Values in parentheses correspond to the outermost shell.

The aforementioned data sets were collected from crystals grown under the same conditions with the cryoprotectant being included in the reservoir solution during crystal growth. Both data collections resulted in orthorhombic lattices exhibiting crystal polymorphism. In the case where glycerol was added in the crystallization medium, crystals belonged to space group  $C222_1$ , while in the case of ethylene glycol, crystals belonged to space group  $P2_12_12_1$  taking into account the systematic absences for every space group.

In order to avoid any possible ambiguity regarding the crystal lattice, both data sets were also processed independently, in  $P1$  space group, followed by POINTLESS (Evans, 2006). The proposed space groups for both data sets were those as previously defined.  $C222_1$ , which is in higher symmetry compared to  $P2_12_12_1$ , has eight symmetry operators and  $P2_12_12_1$  has four. In this case where the unit cell size was approximately the same for both data, the number of molecules per asymmetric unit (ASU) should have been doubled for the lower symmetry space group. Indeed, according to Matthews Coefficient (Winn et al., 2011), for  $C222_1$  there

were two molecules of DSCAM Ig1-Ig4 in the ASU and for P<sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub> there were four molecules in the ASU (Table R-9). In addition, in order to investigate the presence of translational pseudosymmetry and twinning the programs SFCHECK (Vaguine et al., 1999) and CTRUNCATE (French and Wilson, 1978) from CCP4 suite (Winn et al., 2011) were used, respectively. The results showed that no pseudosymmetry or twinning was detected for either space groups (Figure R-32).

Table R-9. Matthews coefficient as calculated by CCP4 suite for DSCAM Ig1-Ig4 WT.

Spacegroup C222<sub>1</sub>, MW: 48 kDa

Nmol/asym	Matthews coef.	% solvent	P (tot)
2	<b>2.61</b>	<b>52.94</b>	<b>0.97</b>
3	1.74	29.41	0.02

Spacegroup P<sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub>, MW: 48 kDa

Nmol/asym	Matthews coef.	% solvent	P (tot)
4	<b>2.98</b>	<b>58.79</b>	<b>0.20</b>
5	2.39	48.49	0.53

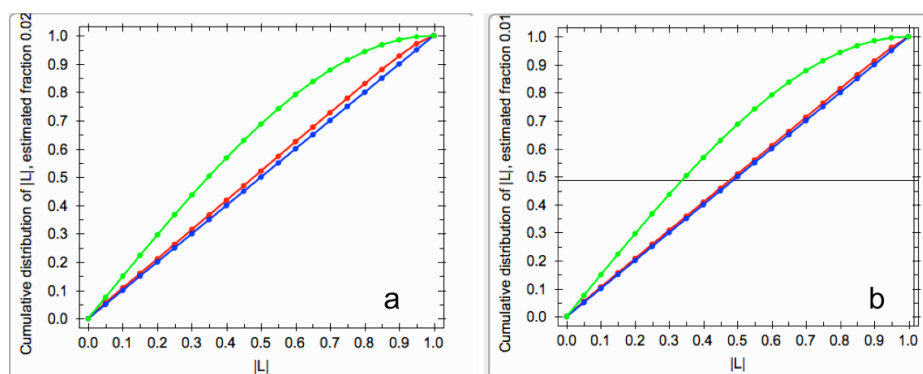


Figure R-32. ILI test for twinning for data collected in space groups (a) C222<sub>1</sub> and (b) P<sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub> using CTRUNCATE (CCP4 suite). The L statistic was 0.479 for C222<sub>1</sub> and 0.491 for P<sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub> (for untwinned data the statistic is 0.5 and for perfect twin is 0.375) (Padilla and Yeates, 2003). The blue and green lines represent theoretical untwinned and perfectly twinned data, respectively. The red curve represents the experimental data indicating that no twinning was detected for either space groups.

In order to determine the 3D structure of human DSCAM Ig1-Ig4, Dscam Ig1-Ig4 from *D. melanogaster* (PDB entry 2V5M, 32 % sequence identity) (Meijers et al., 2007) was used as a search model for both space groups. A polyalanine model, excluding the solvent and glycans, was prepared (Chainsaw, CCP4 suite) and molecular replacement (MR) was conducted with MOLREP (Vagin and Teplyakov, 1997) and Phaser (McCoy et al., 2007) programs, applying standard protocols as implemented in CCP4 suite. The results showed that no solution was found.

Additional search models were tested including the initial polyalanine model with loop regions trimmed (e.g. the 5-residue-linker between Ig2-Ig3 domains), as well as individual ensembles comprising domains Ig1:Ig2 and Ig3:Ig4. An ensemble of Ig1:Ig4 domains was also examined as a search model, since previously determined structures, bearing a horseshoe shape, showed that domains Ig1 and Ig4 were involved in intramolecular interactions forming a more stable substructure (§3.1.6); however, the results obtained from MR, using Phaser, indicated that no solution was found (Table R-10). An automated molecular replacement pipeline, BALBES, was also used towards the structure solution (Long et al., 2008). For space groups C222<sub>1</sub> and P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>, single solutions were found (with 54.1 % and 38.9 % probability, respectively), which were further subjected to refinement (Murshudov et al., 1997, Murshudov et al., 2011) against the experimental data but no convergence to true structure was achieved. The results are summarized in Table R-11.

**Table R-10.** Results obtained from MR using Phaser for DSCAM Ig1-Ig4 WT. Dscam Ig1-Ig4 (PDB entry 2v5m) from *Drosophila* was used as a search model along with various combinations of its Ig domains.

Search models for MR with Phaser	C222 <sub>1</sub>		P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	
	RF LLG/ Z-score	TF LLG/ Z-score	RF LLG/ Z-score	TF LLG/ Z-score
Ig1-Ig4*	21.3/3.87	33.12/6.6	16.5/3.0	19.0/3.3
Ensemble 1: Ig1:Ig2	21.2/4.0	31.35/5.6	-39.8/4.0	-311.9/6.1
Ensemble 2: Ig3:Ig4	11.2/4.1	9.1/5.1	-21.5/3.6	-217.1/6.2
Ig1:Ig4	22.8/4.0	27.9/4.9	-16.2/3.9	-266.2/5.8
Ig1	21.4/3.7	18.2/4.7	21.8/3.4	-13.2/5.7
Ig2:Ig3	6.1/3.6	2.7/4.9	-21.5/4.0	-290.5/4.6

\* Ig1-Ig4, polyalanine model.

RF: Rotation Function, TF: Translation Function, LLG: Log Likelihood Gain.

**Table R-11.** Refinement statistics for DSCAM Ig1-Ig4 WT by BALBES\*.

	C222 <sub>1</sub>		P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	
	Initial	Final	Initial	Final
<b>R</b>	0.5510	0.5340	0.5650	0.5150
<b>R<sub>free</sub></b>	0.5440	0.5250	0.5550	0.5580
<b>Q factor</b>	0.454		0.407	

\* The best solution suggested by BALBES for C222<sub>1</sub>, was based on the search model of the N-terminal fragment from Human TAG-1 comprising 2 Ig domains (PDB entry 2OM5) (Mörtl et al., 2007) and for P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> the structure from Dscam Ig1-Ig4 (PDB entry 2v5m) (Meijers et al., 2007) was used as a search model, providing the best solution.

In parallel, efforts were directed towards obtaining experimental phases with heavy atoms derivatized crystals, using single-wavelength anomalous dispersion (SAD). Diffraction data were collected at beamline P13 (PETRA III, EMBL-Hamburg, DESY, Germany) at 100 K from DSCAM Ig1-Ig4 WT crystals soaked in heavy atom solutions. The concentration of the heavy atoms ranged from 5 mM to 100 mM and different soaking periods were tested (10min - 24h). The majority of the crystals soaked with Gadolinium (Gd-HPDO3A) were immediately dissolved and crystals soaked with Lead (Pb(CH<sub>3</sub>COO)<sub>2</sub>·3H<sub>2</sub>O) showed cracks that prevented further indexing and characterization during data collection. To identify the optimum conditions, the heavy atom concentration and soaking periods were decreased using larger crystals for the experiments. The results obtained showed that crystals diffracted to ~5 Å resolution. Soaking with Ytterbium (YbCl<sub>3</sub>·H<sub>2</sub>O) and Platinum (K<sub>2</sub>PtCl<sub>4</sub>) complexes enabled data collection to 3.3 Å resolution at the Yb L(III) and Pt L(III) absorption edges (with wavelengths 1.386 Å and 1.072 Å, respectively) but no binding was observed as it was indicated by the poor anomalous signal recorded (Tables R-12 & R-13).

Table R-12. Diffraction data statistics of DSCAM Ig1-Ig4 WT soaked with YbCl<sub>3</sub>·H<sub>2</sub>O.

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR expected	COMPARED I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano		
7.98	14638	2063	2067	99.8%	6.9%	6.6%	14635	24.95	7.4%	99.7*	57*	1.578	865
5.65	26584	3706	3711	99.9%	16.5%	14.6%	26581	12.26	17.8%	99.0*	11*	0.968	1683
4.61	34036	4839	4850	99.8%	25.8%	21.6%	34035	8.99	27.8%	97.5*	2	0.925	2250
3.99	41568	5734	5742	99.9%	42.6%	37.1%	41560	5.66	45.9%	94.4*	2	0.874	2697
3.57	45087	6455	6458	100.0%	91.5%	90.1%	45083	2.41	98.9%	79.3*	4	0.781	3059
3.26	53154	7161	7162	100.0%	160.7%	171.5%	53140	1.30	172.8%	56.1*	-7	0.661	3400
3.02	54626	7809	7812	100.0%	318.9%	358.3%	54604	0.61	344.5%	30.2*	-2	0.611	3718
2.82	61346	8354	8354	100.0%	560.9%	641.3%	61316	0.35	603.3%	13.8*	-4	0.547	3988
2.66	39772	7586	8897	85.3%	681.7%	789.9%	39059	0.20	752.1%	5.8	0	0.531	3234

Table R-13. Diffraction data statistics of DSCAM Ig1-Ig4 WT soaked with K<sub>2</sub>PtCl<sub>4</sub>.

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR expected	COMPARED I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano		
8.78	5749	1528	1541	99.2%	6.1%	6.3%	5735	17.02	7.1%	99.3*	-2	0.760	625
6.23	10499	2750	2757	99.7%	10.0%	10.1%	10480	10.91	11.6%	99.0*	-3	0.764	1218
5.09	13412	3546	3555	99.7%	15.6%	15.1%	13384	7.94	18.2%	97.1*	0	0.838	1607
4.41	16251	4216	4224	99.8%	19.6%	18.3%	16239	7.08	22.8%	96.0*	6	0.885	1960
3.95	17448	4768	4782	99.7%	29.9%	28.4%	17415	4.91	35.1%	91.6*	4	0.851	2221
3.61	20671	5282	5291	99.8%	57.2%	55.8%	20658	2.93	66.3%	78.9*	4	0.813	2492
3.34	21754	5661	5669	99.9%	95.8%	98.2%	21721	1.74	111.4%	51.4*	3	0.764	2672
3.12	23483	6146	6154	99.9%	159.3%	168.1%	23449	0.99	185.3%	23.5*	-2	0.683	2913
2.95	20033	5615	6585	85.3%	221.0%	238.0%	19561	0.63	257.8%	14.6*	0	0.649	2372
total	149300	39512	40558	97.4%	31.9%	32.1%	148642	4.36	37.2%	97.1*	2	0.773	18080



Most recently, additional data were collected at 100 K from a single crystal of DSCAM Ig1-Ig4 WT that was previously grown in 13 % PEG 3350, 2 % PGA-LM and 0.1 cacodylate buffer pH 6.5 (Figure R-30d) at beamline P13 (PETRA III, EMBL-Hamburg, DESY, Germany). Prior to data collection, the crystal was flash-soaked into a derivative cryo-solution, containing the precipitant agents (with 5 % PEG increase), 10 % (v/v) glycerol and 1mM K<sub>2</sub>PtCl<sub>4</sub>. The crystal diffracted to 2.35 Å resolution and a complete data set was collected. X-ray data were processed with XDS package (Kabsch, 2010), followed by AIMLESS (Evans, 2006). Diffraction data analysis showed that the crystal symmetry was consistent with space group C222<sub>1</sub> (C-centered orthorhombic lattice), with unit-cell dimensions a= 55.0 Å, b= 146.1 Å, c= 231.5 Å,  $\alpha=\beta=\gamma= 90.0^\circ$ . Data collection statistics for DSCAM Ig1-Ig4 are summarized in Table R-14.

#### **i. Structure determination and refinement of DSCAM Ig1-Ig4 WT**

The 3D structure of DSCAM Ig1-Ig4 was determined at 2.35 Å resolution by Dr. Isabel Bento (Schneider Group) with molecular replacement using MOLREP (Vagin and Teplyakov, 1997). The search model used was generated by the Rosetta Comparative Modeling server (Kim et al., 2004; Song et al., 2013), based on the crystal structure of neurofascin from *H. sapiens* (PDB entry 3P3Y, chain A, 23 % sequence identity, Figure R-26 in green) (Liu et al., 2011). Superposition of DSCAM (monomer A) with neurofascin (monomer A) over C<sub>α</sub> atoms (330 residues), gave an r.m.s.d of 3.7 Å, as calculated with the Lsqkab program (CCP4 suite) (Kabsch, 1976). The correct solution had a contrast of 3.27 and the translation factor contrast (TF/sig) was 8.91. The model was then subjected to ten rounds of rigid body refinement, with isotropic B-factors, using REFMAC (Murshudov et al., 1997, Murshudov et al., 2011) considering each monomer as one rigid domain (starting *R* and *R*<sub>free</sub> values from rigid body refinement: 49.1 % and 50.5 %, respectively). Alternate cycles of restrained refinement with non-crystallographic symmetry (NCS) restraints were also performed, followed by manual building of the model with COOT (Emsley and Cowtan, 2004). The model was then subjected to further rounds of restrained refinement followed by phase improvement using Parrot (Cowtan, 2010). Additional alternating cycles of automated model building and refinement were performed with Buccaneer (Cowtan, 2006, 2008) and REFMAC, respectively. The final coordinates of the structure were optimized with Phenix (Adams et al., 2010) and the structure was refined to a final *R*

factor of 20.54 % and a final  $R_{\text{free}}$  of 25.35 %. Refinement statistics and model validation for DSCAM Ig1-Ig4 are presented in [Table R-14](#).

[Table R-14](#). Data collection and refinement statistics of DSCAM Ig1-Ig4 WT on C222<sub>1</sub>.

<b>Data collection and processing statistics</b>	
Experiment	DSCAM Ig1-Ig4 WT
X-ray source	Beamline P13, PETRA III DESY, Germany
Wavelength (Å)	1.071
No of images	1200
Oscillation angle (°)	0.1
Space group	C222 <sub>1</sub>
Unit cell dimensions (Å)	a=55.0, b=146.1, c= 231.5 $\alpha=\beta=\gamma= 90.0^\circ$
Resolution range (Å)	73.05 - 2.35 (2.43 - 2.35) <sup>c</sup>
No of observations	167040 (13416)
No of unique reflections	39162 (3777)
$R_m$ <sup>a</sup>	0.072 (0.39)
Completeness (%)	99.2 (97.4)
$\langle I/\sigma(I) \rangle$ <sup>b</sup>	15.08 (2.85)
CC <sub>1/2</sub>	0.998 (0.296)
Multiplicity	4.3 (3.6)
<b>Refinement statistics and model quality</b>	
Resolution range (Å)	73.0 - 2.35
No of monomers in the ASU	2 (Chain A, B)
No of reflections	39160
Residues included	A: 3-383 B: 2-383
No of protein atoms	A: 3007 B: 3013
No of heteroatoms	201
Solvent molecules	133
Glycerol molecules	2
GlcNAc molecules	4
Matthews coefficient (Å <sup>3</sup> Da <sup>-1</sup> )	2.59
$R_{\text{free}}$ (%) <sup>d</sup>	25.35
$R_{\text{work}}$ (%)	20.54
<b>R.m.s.d in</b>	
Bond lengths (Å)	0.003
Bond angles (°)	0.723
<b>Molprobrity analysis<sup>e</sup></b>	
Ramachandran favoured/outliers/allowed (%)	93.3/0.65/6.01
Rotamer outliers (%)	9.31
B-factor (Å <sup>2</sup> ) (Wilson Plot)	38.82
<b>Average B-factor (Å<sup>2</sup>) for protein residues<sup>f</sup></b>	
Overall	A: 50.7 B: 47.6
Main chain atoms	A: 48.5 B: 45.3
Side chain atoms	A: 53.0 B: 49.8
<b>Average B-factor (Å<sup>2</sup>) for heteroatoms<sup>f</sup></b>	
Water molecules	43.7
Glycerol molecules	50.3
GlcNAc molecules	78.9

<sup>a</sup>  $R_m = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$ , where  $I_i(hkl)$  is the intensity of a reflection and

$\langle I(hkl) \rangle$  is the mean intensity of all  $i$  symmetry-related reflections.

<sup>b</sup>  $\sigma(I)$  is the standard deviation of  $I$ .

<sup>c</sup> Values in parentheses correspond to the outermost shell.

<sup>d</sup>  $R_{\text{free}}$  is calculated using 5 % of the total reflections that were randomly selected and excluded from refinement.

<sup>e</sup> Molprobit analysis was performed using the Molprobit server (Chen et al., 2010).

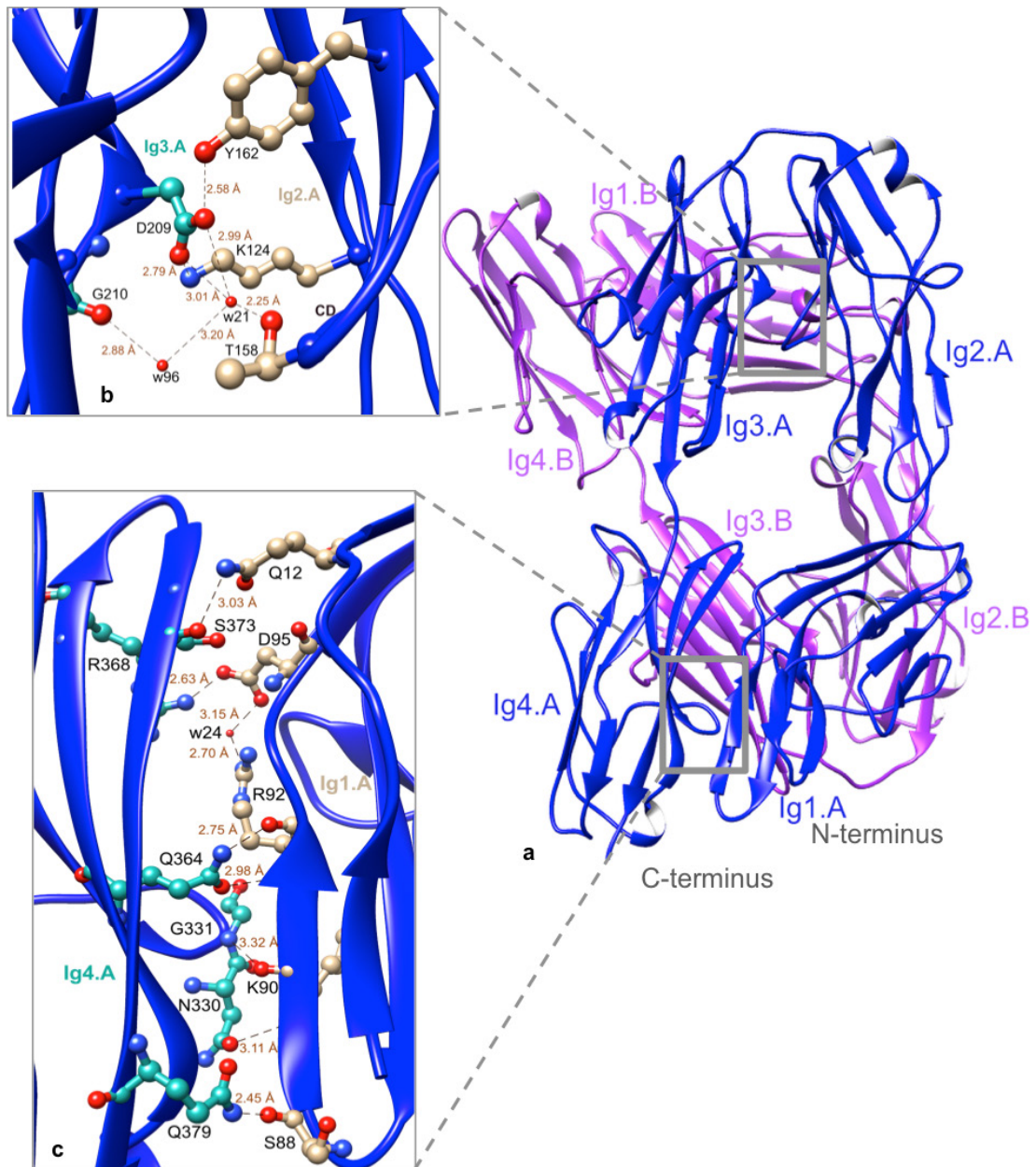
<sup>f</sup> Analysis performed by BAVEGAGE (Winn et al., 2011).

## ii. DSCAM Ig1-Ig4 WT structural analysis

The 3D structure of human DSCAM Ig1-Ig4 revealed the presence of two molecules in the ASU (Monomer A: 381/383 a.a. and Monomer B: 382/383 a.a.) forming a homodimer (Figure R-33). The overall fold of the DSCAM Ig1-Ig4 monomer was similar to the homologous structures from *Drosophila* (Meijers et al., 2007; Li et al., 2016) and to protein structures previously determined bearing a U-shape (horseshoe-shape) configuration (Figure R-26). Each monomer consisted of four Ig-like domains (Ig1: Ser3-Leu102, Ig2: Arg103-Ser198, Ig3: Ala204-Lys291, Ig4: Gln292-Glu383) with a five-residue linker (a.a.:199-203) connecting domains Ig2 and Ig3 (Figure S1). The secondary structure elements of each Ig domain were formed by two  $\beta$ -sheets of antiparallel  $\beta$ -strands, linked with one disulphide bond and adopted the characteristic Ig-fold of the C2-set domain that resembled the antibody constant domain (Fowler and Clarke, 2001). Visual inspection of the electron-density maps during the final stages of refinement revealed additional density at Asparagine residues of monomer A (Asn58) and monomer B (Asn8 and Asn58) that was attributed to GlcNAc molecules, as a result of the glycosylation. More specifically, one GlcNAc molecule was covalently bonded to Asn58.A, one to Asn8.B and two molecules to Asn58.B. The GlcNAc molecules observed were part of the glycosylation chain and were consistent with the prediction results obtained using the NetNGlyc 1.0 Server (Gupta and Brunak, 2002) for DSCAM, based on the Asn-X-Ser/Thr motif for the presence of Asparagine *N*-glycosylation sites. In addition, two glycerol molecules were also included in the structure, originating from the cryoprotectant used, as it was indicated by the difference Fourier maps ( $F_o - F_c$ ) at 3.0 sigma contour level.

*Analysis of the intraface (Ig1.A:Ig4.A and Ig2.A:Ig3.A)*

As it was previously described (§3.1.6), the residues involved in the stabilization of the Ig1:Ig4 intraface (for each monomer) were conserved among structural homologous proteins; however, a sequence comparison of DSCAM with Dscam structures from *Drosophila*, showed that although the amino acids were conserved in human DSCAM as well (Figure R-25), structural comparison revealed that there were differences among the residues contributing to this substructure formation. More specifically, Ser88 and Lys90 from domain Ig1.A, interacted with Gln379 and Asn330-Gly331 from Ig4.A, respectively. In addition the structure was stabilized through a hydrogen-bond network formed among Asp95, Arg368, Arg92 and Gln364 residues of the same monomer (Figure R-33c). The horseshoe-shape structure was further stabilized with intramolecular hydrogen bonds formed among Ig2.A and Ig3.A domains and solvent molecules acting as intermediates. Two residues extending from the CD-loop of domain Ig2 (Thr158 and Lys124) interacted directly with two water molecules and then with Asp209 and Gly210 from Ig3 domain. Although the solvent accessible surface area for Ig2:Ig3 was smaller than for Ig1:Ig4 one, it made a substantial contribution to the maintenance of the horseshoe-shape (Figure R-33b).



**Figure R-33.** The crystal structure of human DSCAM Ig1-Ig4 determined at 2.35 Å resolution. **a.** Ribbon representation of the structure with two molecules of DSCAM forming a homodimer in the ASU (monomer A, shown in blue and monomer B, shown in purple). **b.** Expanded view of the intraface formed between Ig2 (shown in beige) and Ig3 (shown in cyan) of the same chain. **c.** Expanded view of the intraface formed between Ig1 (shown in beige) and Ig4 (shown in cyan) of the same chain. The amino acids forming the interactions are presented in ball and stick with oxygen in red, nitrogen in blue, carbon in beige (for Ig1 and Ig2) and cyan (for Ig3 and Ig4). Water molecules (w) are depicted as red spheres. The molecular graphics representation was performed with UCSF CHIMERA (Pettersen et al., 2004).

### *Analysis of the dimeric interface in the ASU (monomer A:monomer B)*

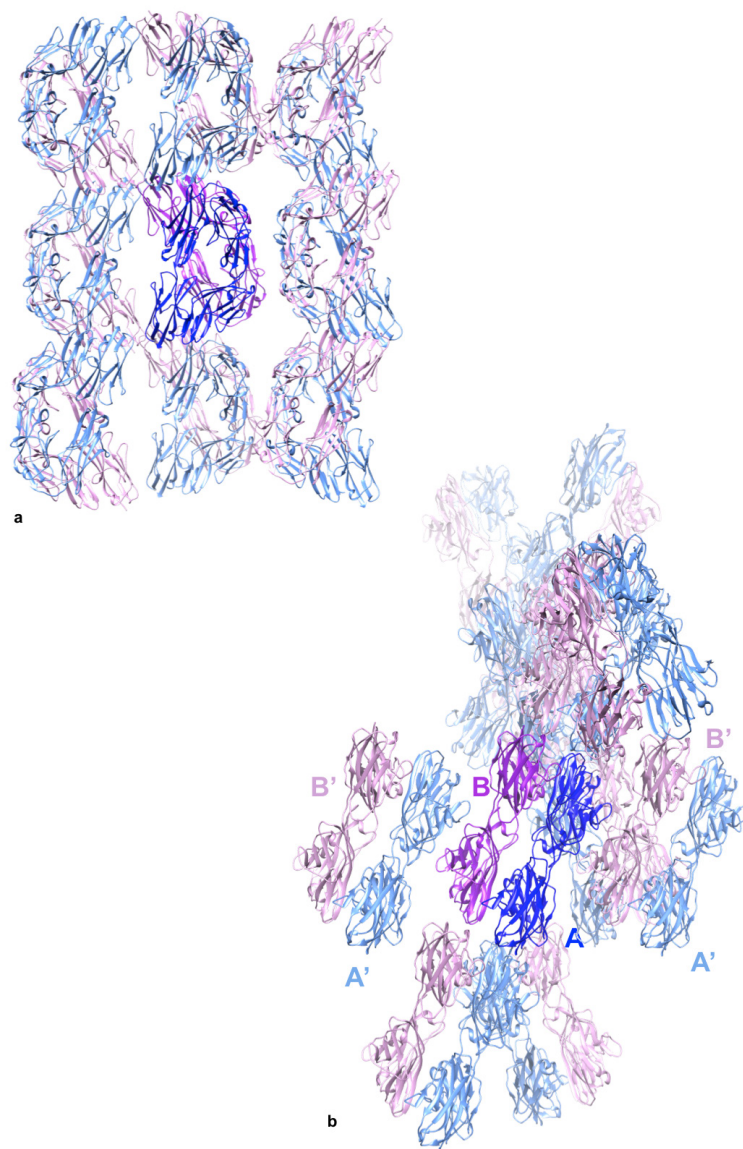
The interface formed between monomer A and monomer B (Figure R-33a, depicted in blue for monomer A and purple for monomer B), was different from the one observed in Dscam structures in *Drosophila* and it was sustained by a hydrogen network involving all Ig domains from both monomers. More specifically, the dimer was stabilized by a salt bridge between Arg287 (Ig3.A) and Glu324 (Ig4.B) and direct hydrogen bonds among residues belonging to Ig1 and Ig2 domains. Residues from Ig1.A domain, Phe71 and Thr73 interacted with Arg109 and Glu111, respectively, from Ig2.B. In addition, residues from domain Ig1.A (Thr42 and Glu44) formed hydrogen bonds with residues from Ig3.B (Lys283 and Ser206). Three residues, Ile207.A, Thr281.B and Asn325.B formed water-mediated hydrogen bond interactions with Asn78.B, Gly43.A and Glu274.A, respectively. The analysis also revealed that the dimer was not stabilized by residues directly involved in hydrophobic interactions.

In order to investigate whether the interface formed was of biological significance or a result of crystal contacts, the buried surface area (BSA) of the interface was calculated using the PISA server (Krissinel and Henrick, 2007). The BSA was found to be 1080 Å<sup>2</sup>, which was below the threshold value of 1200 Å<sup>2</sup>, that was proposed by Janin and co-workers for the formation of specific contact interactions (Janin and Rodier, 1995; Janin, 1997). In addition, the analysis of this protein interface, performed by PISA server, revealed the absence of specific interactions that could result in the formation of a quaternary structure. However, only the calculation of the interface size is not sufficient to distinguish biological dimers from crystal packing interfaces and additional parameters were investigated (Bahadur et al., 2004). Therefore, parameters such as the non-polar interface area ( $f_{np}B$ ) that was 514 Å<sup>2</sup> (46.7 %) and the fraction of fully buried atoms ( $f_{bu}$ ), which was 0.14, were calculated using the BioCOMplexes COnTact MAPS server (Vangone et al., 2011) and the PROFACE server (Saha et al., 2006). According to Bahadur and co-workers studies, the average  $f_{np}B$  was found to be 50-70 % and the average  $f_{bu}$  in the interface of well-characterized structures presenting homodimers was around 0.34-0.36. In addition, determination of the shape complementarity value ( $S_c$ ), calculated with the programme SC (supported by CCP4), was also taken into consideration (Lawrence and Colman, 1993). The shape correlation statistic is used to quantify the shape complementarity of two interacting molecular surfaces, by taking into account distances and angles of both molecules involved in the interface. In the DSCAM case,

the  $S_C$  value was 0.51, which was lower than the expected values reported for determined protein-protein interactions or homodimers ( $0.70 \pm 0.08$ ) (Lawrence and Colman, 1993; Kuroda and Gray, 2016). Therefore, the aforementioned criteria indicated that the homodimer formation of DSCAM Ig1-Ig4 in the ASU, was a result of the crystal packing.

#### *Structural comparison with Dscam Ig1-Ig4 in Drosophila (Ig2.A:Ig2.B' & Ig3.A:Ig3.B')*

Dscam in *Drosophila* (Isoform 1.34, PDB entry 2v5m) consists of 388 residues, has one molecule in the ASU and the biological unit, which is known to be a dimer, is formed with a symmetry related monomer. Superposition of DSCAM (monomer A) with Dscam over  $C_\alpha$  atoms (360 residues), gave an r.m.s.d of 2.7 Å. The rest of Dscam structures corresponding to different isoforms, determined by Li and co-workers (Li et al., 2016) form homodimers with 2 molecules per ASU and the r.m.s.d. values vary from 2.3 to 2.5 Å. The dimeric interface of Dscam, necessary for the repulsion mechanism, is formed due to specific interactions among domains Ig2 and Ig3 from monomer A with their corresponding domains from monomer B (Ig2, Ig3) (Figure R-35a). These residues are either identical or similar among the different Dscam isoforms (Figure S1, highlighted in purple); however they are not conserved in human DSCAM. Since the DSCAM interface formed in the ASU was different from the one observed in *Drosophila*, the contacts of symmetry related molecules of DSCAM that were generated, were investigated to examine whether any of those dimers were mimicking the Dscam interface (Figure R-34a). The symmetry related molecules showed that an interface was indeed formed between monomer B and its symmetric one monomer A' (or between monomer A and B', respectively) (Figure R-34b) that resembles the one formed in Dscam Ig1-Ig4, involving the same domains.

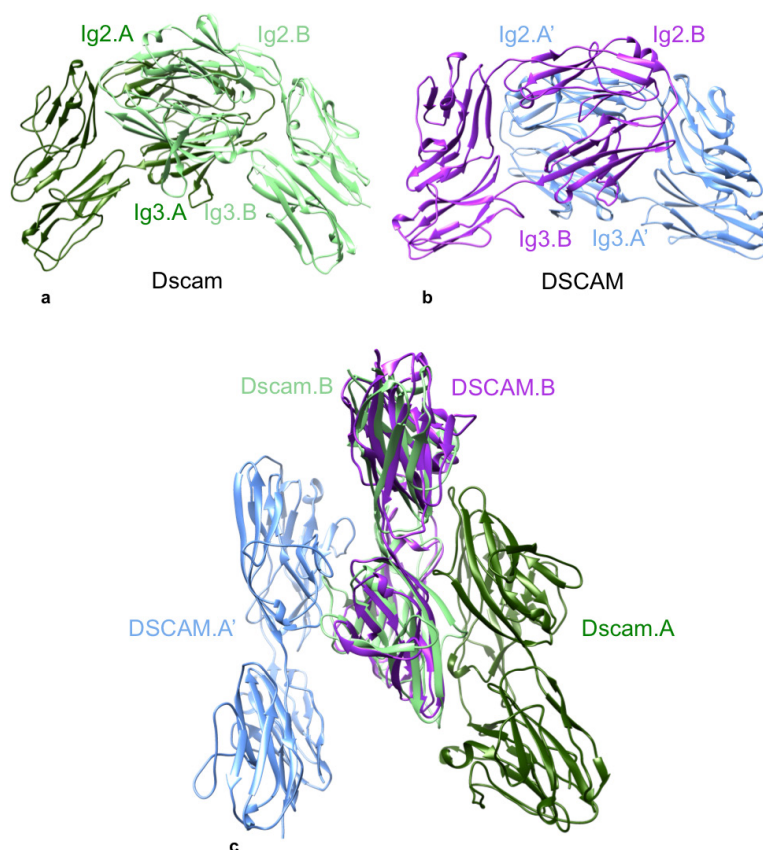


**Figure R-34.** Ribbon representation of DSCAM Ig1-Ig4 molecules in the crystal. **a.** DSCAM molecules in the ASU are depicted in blue for monomer A and purple for monomer B. The symmetry related molecules are shown in cornflower blue and plum for monomers A' and B', respectively. **b.** Rotation of image (a) by 90° along the b axis. A dimer interface is formed between monomer B (purple) and its symmetric A' (cornflower blue) and monomer A (blue) and its symmetric B' (plum), involving interactions among domains Ig2:Ig2' and Ig3:Ig3'. The molecular graphics representation was performed with UCSF CHIMERA (Pettersen et al., 2004).

Despite the structural similarity of human DSCAM monomer with the *Drosophila* one, the aforementioned crystal lattice configuration of the dimer exhibited many differences (Figure 35). More specifically, the orientation of monomer B in DSCAM (Figure 35b, in purple) was different from the corresponding one in Dscam (Figure 35a, in green) and the BSA formed between the monomers was very small compared to the one in *Drosophila*. A closer look in the residues involved in the interactions of these domains in DSCAM, revealed that the amino acids varied from the homologous



structures. In addition, the small number of interactions formed and the small surface of the interface (Table R-15), as calculated by PISA server, did not lead to a biological form of the protein and this assembly, was a result of the crystal packing. In addition, analysis of the rest of the interfaces among the dimer of DSCAM and the symmetrical molecules in the vicinity, suggested that they were also formed due to crystal contacts.



**Figure R-35.** **a.** Ribbon representation of Dscam in *Drosophila* in the ASU (PDB entry 4X83) (Li et al., 2016) (monomers A and B are shown in green and light green, respectively) **b.** Ribbon representation of human DSCAM Ig1-Ig4 structure with a symmetry related molecule (monomers B and A' are shown in purple and cornflower blue, respectively). **c.** Close-up view of figure R-34.b where Dscam dimer from *Drosophila* has been superposed on to human DSCAM for comparison of the interfaces. The interfaces formed involve the same domains in each case; however the amino acids of DSCAM are fewer and different and the BSA is very small compared to the one from *Drosophila*. Superposition was conducted with Lsqkab (CCP4 suite) (Kabsch, 1976) with r.m.s.d. over  $C_{\alpha}$  atoms (352 residues) being 2.3 Å. The molecular graphics representation was performed with UCSF CHIMERA (Pettersen et al., 2004).

**Table R-15.** Buried surface area and number of residues (Nres) involved in interdomain interactions of Dscam Ig1-Ig4 homodimers (Ig2:Ig2 and Ig3:Ig3) in comparison to DSCAM interface involving the same domains (Ig2:Ig2' and Ig3:Ig3').

<b>PDB entry</b>	<b>Human DSCAM</b>	<b>4X83</b>	<b>2v5m</b>	<b>4X9B</b>	<b>4X9F</b>	<b>4X8X</b>	<b>4X9H</b>	<b>4XB8</b>	<b>4X9G</b>	<b>4XB7</b>
<b>BSA (Å<sup>2</sup>)</b>	350	1174	1795	1496	1723	1323	2193	1646	1642	2171
<b>Nres</b>	13	40	50	50	48	41	62	54	53	64
<b>Resol. (Å)</b>	2.35	1.90	1.95	2.20	2.35	2.50	2.95	3.20	3.40	4.0

Dscam Ig1-Ig4 structure with PDB entry 2v5m was determined by (Meijers et al., 2007) and the rest of Dscam structures representing different isoforms by (Li et al., 2016).

## 3.2 Discussion

Previous studies have shown that Dscam in insects is expressed in ~38,000 isoforms with ~19,000 distinct extracellular domains (Armitage et al., 2012). There is evidence that this molecular diversity is essential for neuron recognition and self-avoidance among branches, arising from the same cell (Chen et al., 2006; He et al., 2014). This seems to become possible through specific homophilic interactions among the same Dscam isoforms (homodimerization) (Schmucker and Chen, 2009; Wojtowicz et al., 2004, 2007). On the contrary, mammalian DSCAM does not exhibit isoform variability. Furthermore, animal studies conducted in mice proposed that DSCAM is mainly implicated in heterophilic interactions with guidance cues like netrin-1, necessary for neuronal connectivity and commissural axon guidance. These *in vivo* assays demonstrated that the DSCAM domains involved in netrin-1 binding are Ig7, Ig8 and Ig9 (Ly et al., 2008). However, in insects according to structural studies, domain Ig7 is required for Dscam dimerization (Wojtowicz et al., 2007). Therefore, one could suggest that the mechanism dictating DSCAM homophilic binding might be different or in case it is conserved, it might be disrupted by netrin-1 binding and as a result, the interactions observed in insects might not be formed in the same way in vertebrates.


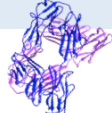
With the aim to explore the mechanism underlying DSCAM homodimerization, biophysical and structural studies of human DSCAM such as ITC, SAXS, EM and X-ray protein Crystallography, were performed. Two hypotheses were examined; i) DSCAM homophilic interactions in vertebrates are formed without the involvement of the same domains as in insects, ii) Binding of netrin-1 interferes with DSCAM homodimerization by disrupting the dimer interactions. In order to investigate these hypotheses and to assess how netrin is implicated in heterophilic interactions with DSCAM, the aforementioned studies included netrin<sub>VIV</sub> in association with DSCAM as well.

To examine the above hypotheses, two constructs of human DSCAM ectodomain were designed, comprising eight and nine Ig domains (DSCAM Ig1-Ig8 and DSCAM Ig1-Ig9), respectively. Both constructs were expressed in HEK 293S and 293T cells. DSCAM Ig1-Ig8 elution profile after purification showed that it was predominantly a dimer in solution, while the rest was in a monomeric state (Figure R-3). Moreover, introduction of the 9<sup>th</sup> domain to the construct, further stabilized the dimer interactions, suggesting that this domain was also involved in DSCAM

homodimerization process. Indeed, DSCAM Ig1-Ig9 was always eluted as a dimer (in monodisperse population) (Figure R-5), in agreement with previous studies reported in the literature for human DSCAM.

In addition, a smaller construct was designed (DSCAM Ig1-Ig4) to investigate by structural methods the plausible formation of a more rigid configuration. Determination of DSCAM Ig1-Ig4 3D crystal structure would not provide a thorough answer on the mechanism under which DSCAM domains are implicated in homodimerization; however, it would facilitate spotting the differences of the first four Ig domains between humans and insects (in comparison to the 3D structures of *Drosophila* of the same construct (Meijers et al., 2007; Li et al., 2016)). DSCAM Ig1-Ig4 was expressed both in the WT and GnTI<sup>-</sup> form and the results obtained, from different methods, indicated that it was a stable monomer in solution (Figure R-28). The results from all DSCAM constructs are summarized in Table D-1.

Table D-1. Summary of methods employed and results obtained for human DSCAM.

	DSCAM Ig1-Ig9	DSCAM Ig1-Ig8	DSCAM Ig1-Ig4
<b>Cloning</b>	✓	✓	✓
<b>Expression WT/GnTI<sup>-</sup></b>	✓/✓	✓/✓	✓/✓
<b>Purification WT/GnTI<sup>-</sup></b>	✓/✓	✓/✓	✓/✓
<b>Crystallization WT/GnTI<sup>-</sup></b>	✓/✓	-/✓	✓/✓
<b>X-ray Data Collection WT/GnTI<sup>-</sup></b>	9 Å	6.7 Å	2.7 Å
<b>EM WT/GnTI<sup>-</sup></b>	✓/✓	✓/✓	-
<b>SAXS WT/GnTI<sup>-</sup></b>	✓	✓/✓	✓
<b>Oligomerization state in solution WT/GnTI<sup>-</sup></b>	Dimer/Dimer	Dimer&monomer/ Dimer&monomer	Monomer/Monomer
<b>EM 2D particle analysis</b>		-	-
<b>Crystal Structure of WT at 2.35 Å</b>	-	-	

More specifically, SAXS measurements were performed for all DSCAM constructs as well as for DSCAM Ig1-Ig9:netrin<sub>VIV</sub> using a range of concentrations. For DSCAM Ig1-Ig4 all models generated had good correlation with the experimental data, demonstrating a horseshoe domain arrangement resembling Dscam in *Drosophila* (Figure R-29), but in a monomeric state. In contrast to DSCAM Ig1-Ig4, DSCAM Ig1-Ig9 was shown to form always dimers in solution (monodisperse population), whereas DSCAM Ig1-Ig8 was found both in monomeric and dimeric states, with the dimeric state being the prevalent one. More specifically, DSCAM Ig1-Ig9 dimer was depicted as an ensemble of models, revealing the overall shape of the molecule. This shape indicated that the first four Ig domains were not part of the dimerization interface forming an independent core, away from the main dimeric body (Figure R-10). All the proposed model structures generated by SAXS, despite the various configurations adopted by the domains involved in the dimeric interface, shared two rigid T-shaped branches lying at the periphery of this interface bilaterally. These extensions corresponded to the Ig1-Ig4 domains and were fitting optimally the experimental data.

In addition, DSCAM Ig1-Ig8 averaged dimeric model from SAXS was also compared with Dscam Ig1-Ig8 crystal structure from *Drosophila* (Sawaya et al., 2008). The results showed that human DSCAM was poorly correlated with Dscam in *Drosophila* since the computed theoretical scattering curve extracted from the crystallographic structure, could not fit the SAXS experimental DSCAM data. Superposition of the two models showed that there were large differences in the radius of gyration suggesting that human DSCAM had a larger and possibly more flexible dimer assembly than the one occurring in *Drosophila* (Figure R-9). The observations extracted from the larger constructs and the fact that DSCAM Ig1-Ig4 does not dimerize in solution, support the working hypothesis that DSCAM homodimerization follows, probably, a different mechanism in humans, than the one observed in *Drosophila* (where Ig1-Ig4 was shown to be a dimer) without the involvement of the four *N*-terminal Ig domains in the dimer interface. Nevertheless, direct comparison of DSCAM in solution with Dscam crystal structure in *Drosophila*, might not provide an immediate answer, since Dscam molecules might be involved in crystal packing interactions imposed by the symmetry of the crystal that might not be present in DSCAM when in solution.

SAXS measurements of DSCAM Ig1-Ig9 with netrin<sub>VIV</sub> (1:1 molar ratio) were compared to the individual plots acquired from netrin<sub>VIV</sub> (Finci et al., 2014) and

DSCAM when alone in solution. The samples in solution did not seem to be engaged in complex formation, as supported by the fact that oligomer analysis showed a mixture of dimeric DSCAM (70 %), monomeric netrin<sub>VIV</sub> (17 %) and monomeric DSCAM (13 %) in the population, with  $\chi^2$  being 1.4. Alternatively, the two proteins could interact and analysis was performed by constructing rigid-body models of the complex. These models fit slightly better the experimental data ( $\chi^2 \sim 1.2$ ) and some structures seemed more plausible based on the statistical analysis performed for ranking the structures energetically (Figure R-22). However, in order to resolve the discrepancy between the two different results reported from SAXS measurements, additional evidence would be required regarding the stoichiometry and binding affinity upon complex formation (e.g. data from binding assays, such as ITC or SPR).

Negative staining electron microscopy experiments were performed with DSCAM Ig1-Ig8, Ig1-Ig9 and DSCAM Ig1-Ig9:netrin<sub>VIV</sub>. In the case of DSCAM Ig1-Ig8, the sample showed increased heterogeneity after negative staining, preventing further single-particle analysis. This could be attributed to its polydispersity observed in solution (mixture of monomers & dimers). On the contrary, DSCAM Ig1-Ig9 raw images were less heterogeneous, as it was anticipated, and indicated that the population of DSCAM was primarily in monomeric form (even though it was known to be a stable dimer in solution). Taking into account these observations, single-particle analysis and classification were performed with the longer construct in order to produce two-dimensional projections of the protein. The results demonstrated that DSCAM Ig1-Ig9 adopted different conformations where the first four Ig domains (Ig1-Ig4) formed a rigid horseshoe core, present in most of the different classes. The Ig1-Ig4 arrangement confirmed the results obtained from SAXS regarding the shape of the domains, showing that this horseshoe is conserved among the species. The 2D projections also revealed that DSCAM Ig1-Ig9 exhibited a variety of orientations that could be attributed to different angular distributions of the molecule on the carbon support film used for negative staining (Frank, 2006).

Domains (Ig6-Ig9) followed a linear arrangement, which appeared to move either away or leaning towards the horseshoe configuration through hinges potentially located in the tandem of domains Ig5-Ig8 (Figure R-12). With the aim to further investigate the location of the hinges, sequence alignment of DSCAM Ig1-Ig9 was performed with protein receptors bearing pliant hinge regions and similar domain organization, folding and size. Sequence comparison showed that only the residues in close proximity to the hinge regions were conserved, whereas the amino acids

belonged to hinges were being rather diverse. These findings confirmed previous studies, where it was suggested that only the sequence of the protein is not sufficient for predicting the hinge areas and the use of computational tools was proposed for producing concrete results (Flores et al., 2007, 2008; Shamsuddin et al., 2013).

In contrast to the results obtained by SAXS where DSCAM Ig1-Ig9 was a dimer in solution, the majority of the EM projections showed a monomer of DSCAM. This observation potentially provides some evidence for the dissociation constant of DSCAM, which might be below the dilution range used for EM negative staining (~ 5-10  $\mu\text{g/ml}$ ), due to weak interactions during homodimer formation. Dimer dissociation, could also be attributed to the stain used (uranyl formate), since although it enhanced the scattering contrast compared to other stains, it might have interfered with DSCAM dimer formation due to its low pH (4.5-5.2) (Booth et al., 2011). Finally, dehydration of the protein as a result of negative staining could also lead to dissociation of the dimeric complexes depending on their stability.

To ensure that DSCAM homodimers could be observed, efforts were directed towards preparing stable DSCAM homodimers, by crosslinking, prior to negative staining. Despite the variety of buffer solutions and salt concentrations used, the sample remained heterogeneous consisting of a mixture of crosslinked and non-crosslinked dimer populations. Separation of the two populations with standard chromatographic techniques could not be achieved since the oligomeric state of the protein was the same (DSCAM was a dimer before and after crosslinking). The raw images after negative staining further supported this finding, resulting in rather heterogeneous populations, which prevented additional analysis and classification.

Regarding DSCAM and netrin, although previous studies in neurons suggested that DSCAM interacts with netrin-1 in the cell (Liu et al., 2009; Huang et al., 2016), the results obtained *in vitro* by EM after image classification, showed a mixture of DSCAM Ig1-Ig9 monomers and larger complexes (most likely of DSCAM homodimers) under the conditions examined.

More specifically, the abundant averages exhibited that DSCAM Ig1-Ig9 was in monomeric form with the same conformational arrangement as the one adopted in the absence of netrin<sub>VIV</sub>. The projections where DSCAM Ig1-Ig9 appeared to bind to another molecule, suggested that this partner was probably a partial molecule of DSCAM, based on the shape and size of its domains compared to netrin<sub>VIV</sub> (Figure R-23). A potential reason why the horseshoe of the partner molecule was not observed, could be attributed to the orientation of the complexes that might have

obscured its complete observation. Alternatively, one of the horseshoe domains might have been flexible (in different orientations) and could have been averaged out during image classification.

Overall, it can be concluded that the addition of netrin<sub>VIV</sub> in DSCAM solution, induced the observation of dimers. The presence of netrin<sub>VIV</sub> probably changed the orientation of DSCAM on the carbon film, leading to new views where the homodimer could be observed. Although, the visualization of the dimer was a significant finding, in order to resolve the ambiguity regarding the identity of the second molecule, additional experiments were performed, by examining the proteins' interaction in different stoichiometry (in favour of netrin). DSCAM Ig1-Ig9:netrin<sub>VIV</sub> complex was still not clearly detected under the conditions used in this study, resulting in similar projections as in 1:1 molar ratio experiments. This suggests that a number of restraints needs to be considered. The low MW of netrin<sub>VIV</sub> (56 kDa) might have prevented the observation of the protein. In addition, incomplete stain embedding might have also interfered with the complex visualization. The biochemical and biophysical experiments performed in this study support the view that if netrin-1 does bind to DSCAM, this might occur via weak interactions (low affinity complex formation). Furthermore, key processes that take place in the cell at molecular level during DSCAM:netrin-1 complex formation, should also be taken into consideration. For example, complex formation in the cell could be induced by other determinants, such as the orientation of DSCAM receptor in the cell membrane that promotes binding of netrin-1.

In order to determine the binding affinity and stoichiometry between DSCAM and netrin<sub>VIV</sub>, binding assays were performed with ITC using both DSCAM Ig1-Ig8 and Ig1-Ig9 domains with netrin<sub>VIV</sub>. The lack of identifying a common stabilization buffer for ITC measurements for both proteins, prevented drawing any conclusions regarding the complex formation. This was mainly due to the propensity of netrin<sub>VIV</sub> to aggregate, despite the variety of buffers and pH tested. *In vitro* reconstitution of the DSCAM Ig1-Ig9:netrin<sub>VIV</sub> complex was also attempted, by mixing and co-purifying the two components using SEC (in 1:1 and 1:2 molar ratios), but no complex formation was observed under the conditions examined.

In order to obtain more concrete information on the domains implicated in the homodimerization of DSCAM, structural studies with X-ray crystallography were performed using the Ig1-Ig8 and Ig1-Ig9 constructs. Diffracting crystals of DSCAM Ig1-Ig8 were only grown for the GnTI<sup>I</sup> form after excessive crystallization trials,



achieving low-resolution. Standard protocols for optimizing the initial conditions were further applied, but the diffraction quality of the crystals remained poor (6.7 Å resolution) (Table R-2). Protein methylation and the effect of EndoH treatment were also explored, but the resulting crystals did not diffract beyond 6.7 Å resolution. Crystallization of DSCAM Ig1-Ig9 was straightforward for both WT and GnTI<sup>-</sup> variant. This could be attributed to the addition of the Ig9 domain that possibly stabilized the homodimer interactions as indicated by biophysical studies. Yet, crystals only diffracted to 9 Å resolution.

Concerning the smaller construct of DSCAM Ig1-Ig4, crystals were obtained in similar conditions with the Ig1-Ig9 construct, after two rounds of optimization. For the WT form, two complete data sets were collected at 2.7 Å (in C222<sub>1</sub>) and at 3.1 Å resolution (in P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>) (Tables R-7 & R-8), indicating that the more rigid configuration adopted by this shorter form was more prone to crystallize and diffract X-rays to higher resolution, compared to the larger constructs. The primitive orthorhombic space group was only observed when instead of glycerol, ethylene glycol was incorporated in the crystallization conditions during crystal growth. Molecular replacement did not provide a solution when Dscam Ig1-Ig4 from *Drosophila* was used as a search model. This could be explained by the angle that domains Ig1-Ig2 form in Dscam, which might not be present in the human structure. In addition, the sequence homology of Dscam (~30 %) coupled with the resolution that the data were collected, increased the level of difficulty in solving the crystal structure of human DSCAM. Derivatization of the crystals using heavy atoms (Yb and Pt compounds), demonstrated poor anomalous signal not sufficient for phase determination. However, our hypothesis that DSCAM Ig1-Ig4 would form a more stable substructure, as suggested by EM studies (horseshoe configuration) and previously known 3D homologous structures, which would be more prone to crystallize, has been confirmed. The Ig1-Ig4 diffracted to 2.7 Å resolution, which is the highest resolution achieved from all the constructs studied, until recently that X-ray crystallographic data, were collected at 2.35 Å resolution by Dr. Isabel Bento, from previously grown crystals under the same conditions. The three-dimensional structure was determined in C222<sub>1</sub> space group by molecular replacement using the structure of neurofascin as a search model (PDB entry 3P3Y, 23 % sequence identity) (Liu et al., 2011). The structure analysis revealed the presence of a homodimer in the ASU, with each monomer (A and B) consisting of four N-terminal Ig domains that adopted a

horseshoe-shape arrangement in agreement with the results obtained from EM and SAXS studies.

The interactions that DSCAM Ig1-Ig4 dimer formed in the ASU and in the crystal were examined. DSCAM Ig1-Ig4 dimer was stabilized via hydrogen bond interactions involving domains (Ig1-Ig4) from both chains, while additional contacts were formed between each monomer and the symmetry related molecules in the crystal. Analysis of the interfaces among the dimer and the symmetrical molecules in the vicinity (generated for  $C222_1$ ) was performed by PISA. In addition, a set of parameters employed for the characterization of dimeric interfaces of determined protein structures ( $BSA$ ,  $S_C$ ,  $f_{np}B$ ,  $f_{bu}$ ) were calculated for DSCAM Ig1-Ig4 dimer. The overall analysis showed the absence of specific interactions that could result to a homodimer structure and the dimers observed were not biological assemblies of DSCAM Ig1-Ig4.

In addition, DSCAM Ig1-Ig4 dimer was compared to the corresponding one in *Drosophila* (the biological unit of which was a dimer) and as it was anticipated, the domains implicated in dimer formation were different. Furthermore, no interactions between the individual monomers of DSCAM Ig1-Ig4 with the symmetry related molecules in the crystal were found to mimic the ones observed in Dscam Ig1-Ig4.

According to the experimental evidence provided by SLS coupled with FPLC and SAXS studies, DSCAM Ig1-Ig4 was a monomer in solution and it was not part of the dimerization interface of DSCAM Ig1-Ig9. The crystal structure of DSCAM Ig1-Ig4 along with the analysis on the interactions formed, suggested that dimer formation was attributed to the crystal packing interactions in  $C222_1$  space group, confirming that the functional form of the protein is probably monomeric. Despite the conservation of the horseshoe shape in both species, the structural and biophysical data analysis allow to conclude that the mechanism of DSCAM in humans does not follow the one in *Drosophila*, since there are differences on the interactions formed between the two species regarding the first four Ig domains. In addition, the rigidity of DSCAM Ig1-Ig4 3D structure could be exploited as a search model for solving the crystal structure of DSCAM Ig1-Ig8 using the low-resolution data collected at 6.7 Å, so as to resolve the sites of interaction upon dimerization.

## Chapter 4

### CONCLUSIONS AND PERSPECTIVES

With the aim to reveal the domains involved in DSCAM homodimerization in humans and how this mechanism is affected upon netrin-1 binding, biophysical and structural approaches were employed on DSCAM Ig1-Ig8, Ig1-Ig9 and their complexes with netrin<sub>VIV</sub>. In parallel, structural studies of a smaller DSCAM construct involving domains Ig1-Ig4 were performed with the aim to study a more rigid substructure and by determining its three-dimensional structure, to assess the role of these domains in human homodimerization.

The results obtained provide an indication that there is not a universal binding mechanism of the Down Syndrome molecule among all species, in contrast to the common mechanism proposed in the literature until present (Sawaya et al., 2008). DSCAM domains Ig1-Ig4 adopt a rigid horseshoe arrangement (conserved among the species) similar to the one occurring in Dscam in *Drosophila* but they are not part of the dimerization interface, which probably lies among the rest of the domains (Ig5-Ig9). This is supported by the fact that DSCAM Ig1-Ig4 does not dimerize in solution (compared to the longer construct) and that domain Ig9 appears to be involved in DSCAM dimer formation by stabilizing its interactions. The differences between the two species are also indicated in the results obtained from SAXS data in solution (for DSCAM Ig1-Ig8&Ig1-Ig9), where the ensemble of homodimer models has a different arrangement from the one observed in the *Drosophila* crystal structure. In addition, the sequence alignment between the two species (Figure S1), showed that the amino acids involved in the dimer interface of Dscam, are not conserved in the human homologue, supporting the aforementioned hypothesis.

However, to consolidate this suggestion and shed light on the mechanism in vertebrates, further experiments are required. Direct comparison between insects and humans could be facilitated by performing SAXS studies on the corresponding Dscam Ig1-Ig8 and Dscam Ig1-Ig4 constructs in *D. melanogaster*. These would provide evidence for the oligomeric state of Dscam in solution. However, 2D diffusion of DSCAM in the cell membrane should also be taken into consideration, because additional restrains are imposed by the bilipid membrane that affect DSCAM orientation in the cells, in comparison with the orientations that DSCAM might adopt *in vitro* in solution. A clear evidence for improving our understanding of human

DSCAM structure-function relationship would be provided by optimization of DSCAM Ig1-Ig9 crystallization conditions to allow its 3D structure determination.

Electron microscopy studies with negative staining revealed that DSCAM Ig1-Ig9 was primarily observed in a monomeric state both in the absence and in the presence of netrin<sub>VIV</sub>, under the conditions examined in this study. For the first time DSCAM nine Ig domains were clearly visualized with the four *N*-terminal domains forming a horseshoe arrangement. Ig6 to Ig9 domains were shown to adopt a linear configuration that could move either towards or away from the core, around potential hinge axes that divided the molecules into rigid segments. These plausible hinge axes provide an indication of the dynamic motion of DSCAM, supporting its key role in forming interactions with its counterparts (i.e. netrin-1) or with itself. A comparison study of proteins (with previously determined 3D structures) bearing a similar domain organization with DSCAM, showed that the majority of hinge regions were identified close to sites that the proteins were anticipated to host their partners (intermolecular interface). However, sequence alignment of these proteins with DSCAM Ig1-Ig9 did not reveal a conserved motif on the hinge regions. To further investigate the dynamic motions of DSCAM, the 3D elucidation of a subfragment structure comprising domains Ig6-Ig9 or Ig7-Ig9 might be the key for investigating the hinge mechanism. In addition, the use of computational approaches e.g. molecular dynamic simulations (Shamsuddin et al., 2013), might also provide an indication of the mechanism, by prediction of regions that might act as hinges.

In the presence of netrin<sub>VIV</sub>, DSCAM was predominantly depicted as a monomer, after negative staining; however, the dimeric population that was also observed was probably induced by the presence of netrin, resulting in averages where DSCAM Ig1-Ig9 was bound to another partner, not being able to distinguish clearly the identity of the second molecule. To fully visualize either the homodimer or the complex DSCAM Ig1-Ig9:netrin<sub>VIV</sub> and reveal the domains involved in these interactions, further research is required. Preliminary studies have already been performed and the findings have illustrated the parameters that could be further examined and optimized. These include, replacement of the stain used, with alternatives reacting at physiological pH, such as ammonium molybdate (pH range 5-7), or vanadium-based stains (pH range 7-8) (Nanovan). These stains although they produce a lower electron density, could result in the homodimer or heterodimer preservation since they are more stable and mild with biological samples (Tracz et al., 1997). Alternatively, cryo-EM could also be used since with this method the molecules might

be kept intact, leading to complex preservation and shape integrity, despite the low contrast. Recent studies have managed to ameliorate the low signal-to-noise ratio by the use of phase plates (i.e. Volta Phase Plate) resulting in images with higher contrast that even allow visualization of smaller proteins, such as netrin<sub>VIV</sub> (56 kDa) (Danev and Baumeister, 2016). Additional efforts towards elucidating the identity of the second molecule could also be focused on netrin<sub>VIV</sub> labeling, using an electron-dense probe i.e. Nanogold (Nanoprobe Inc). This would enable distinguishing whether netrin<sub>VIV</sub> is part of the dimers observed.

SAXS analysis indicated that a plausible complex between DSCAM Ig1-Ig9 and netrin<sub>VIV</sub> could be formed under the conditions used in the study; however to confirm complex formation and monitor the heterophilic interactions, even at low-range concentration, optimization of the stoichiometry or crosslinking of the two individual components could be essential. To determine stoichiometry and binding affinity between DSCAM Ig1-Ig9 and netrin<sub>VIV</sub>, binding assays with ITC were performed. However, the aggregation of netrin<sub>VIV</sub> prevented the determination of the two proteins affinity. Therefore, methods such as native MS or SPR could be employed, in order to assess the stoichiometry and the mass of the molecular assembly, or the binding affinity, using only small amount of proteins. Alternatively, any direct interaction could be revealed by the use of pull down assays between DSCAM and netrin<sub>VIV</sub> or co-expression of the two components *in vitro* could also result in a soluble protein complex. However, the possibility that the interaction between the two proteins in the cell, might be facilitated by the presence of additional molecules induced by DSCAM expression, as suggested by previous results obtained from cell binding assays (Liu et al., 2009), can not be excluded. Therefore, further research is required to elucidate whether another protein is also essential for DSCAM:netrin-1 complex formation.

The 3D structure determination of DSCAM was rather challenging because DSCAM comprises nine Ig domains and hinge regions. Out of these, only the domains involved in the horseshoe configuration (Ig1-Ig4) presented a rigid structure, while the rest seemed to be rather flexible. The hypothesis that DSCAM Ig1-Ig4 would be more prone to crystallize and diffract to higher resolution was correct and a full data set was collected to 2.7 Å resolution. Most recently, a higher resolution data set was collected, allowing the structure determination of DSCAM Ig1-Ig4 at 2.35 Å resolution. This enabled direct comparison with its homologue in *Drosophila* providing evidence about the participation of these domains in DSCAM dimerization in mammals. The 3D structure of human DSCAM Ig1-Ig4 revealed that its

dimerization profile is different from the one occurring in *Drosophila*. The first four domains of DSCAM formed a dimer in the crystal; however this might be attributed to crystal packing interactions imposed by the crystallization conditions used. This result is consistent with the rest of the studies performed in this research project (in solution), suggesting that DSCAM Ig1-Ig4 biological unit is monomeric.

The present study demonstrates that the underlying mechanism of human DSCAM homophilic interactions differs from Dscam in *Drosophila*; however, this finding requires further investigation, including the 3D structure of DSCAM Ig1-Ig9 in order to elucidate its functional role. Overall, the results obtained contribute to the preliminary characterization of DSCAM and reveal the structural differences of the two species. In addition, the integrated structural biology approaches that were applied, determined the workflow towards understanding the structure-function relationship that dictates the interactions of DSCAM and netrin-1. Deciphering the role of cell receptors in axon guidance in vertebrates requires holistic approaches to dissect the complexity of this biological system.

## REFERENCES

- Abts, A., Schwarz, C.K.W., Tschapek, B., Smits, S.H.J., and Schmitt, L. (2012). Rational and Irrational Approaches to Convince a Protein to Crystallize. *Mod. Asp. Bulk Cryst. Thin Film Prep.* *22*, 497–528.
- Adams, P.D., Afonine, P. V, Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilligere, T.C., and Zwart, P. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* *66*, 213–221.
- Agarwala, K.L., Ganesh, S., Tsutsumi, Y., Suzuki, T., Amano, K., and Yamakawa, K. (2001). Cloning and functional characterization of DSCAML1, a novel DSCAM-like cell adhesion molecule that mediates homophilic intercellular adhesion. *Biochem. Biophys. Res. Commun.* *285*, 760–772.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). *Molecular Biology of the Cell* (Garland Science).
- Andrews, G.L., Tanglao, S., Farmer, W.T., Morin, S., Brotman, S., Berberoglu, M. A., Price, H., Fernandez, G.C., Mastick, G.S., and Kidd, T. (2008). Dscam Guides Embryonic Axons By Netrin-Dependent And Independent Functions. *Development* *135*, 3839–3848.
- Antonarakis, S.E., and Epstein, C.J. (2006). The challenge of Down syndrome. *Trends Mol. Med.* *12*, 473–479.
- Aricescu, A.R., and Owens, R.J. (2013). Expression of recombinant glycoproteins in mammalian cells: towards an integrative approach to structural biology. *Curr. Opin. Struct. Biol.* *23*, 345–356.
- Aricescu, A.R., Lu, W., and Jones, E.Y. (2006). A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr. D. Biol. Crystallogr.* *62*, 1243–1250.
- Armitage, S.A.O., Freiburg, R.Y., Kurtz, J., and Bravo, I.G. (2012). The evolution of Dscam genes across the arthropods. *BMC Evol. Biol.* *12*, 1–15.
- Bahadur, R.P., Chakrabarti, P., Rodier, F., and Janin, J. (2004). A Dissection of Specific and Non-specific Protein-Protein Interfaces. *J. Mol. Biol.* *336*, 943–955.
- Bashaw, G.J., and Klein, R. (2010). Signaling from axon guidance receptors. *Cold Spring Harb. Perspect. Biol.* *2*, 1–16.
- Bell, C.H., Healey, E., van Erp, S., Bishop, B., Tang, C., Gilbert, R.J.C., Aricescu, A.R., Pasterkamp, R.J., and Siebold, C. (2013). Structure of the repulsive guidance molecule (RGM)-neogenin signaling hub. *Science* *341*, 77–80.
- Bergfors, T. (2003). Seeds to crystals. *J. Struct. Biol.* *142*, 66–76.
- Bernhardt, R.R., Nguyen, N., and Kuwada, J.Y. (1992). Growth cone guidance by floor plate cells in the spinal cord of zebrafish embryos. *Neuron* *8*, 869–882.
- Blundell, T.L., and Johnson, L.N. (1976). *Protein crystallography* (Academic Press).
- Bohne-Lang, A., and Von der Lieth, C.W. (2005). GlyProt: In silico glycosylation of proteins. *Nucleic Acids Res.* *33*, W214–W219.
- Boivin, S., Kozak, S., and Meijers, R. (2013). Optimization of protein purification and characterization using Thermofluor screens. *Protein Expr. Purif.* *91*, 192–206.

- Boivin, S., Kozak, S., Rasmussen, G., Nemtanu, I.M., Vieira, V., and Meijers, R. (2015). An integrated pipeline for sample preparation and characterization at the EMBL@PETRA3 synchrotron facilities. *Methods*. *95*, 70–77.
- Booth, D.S., Avila-Sakar, A., and Cheng, Y. (2011). Visualizing Proteins and Macromolecular Complexes by Negative Stain EM: from Grid Preparation to Image Acquisition. *J. Vis. Exp.* *58*, 1–8.
- Bouyain, S., and Watkins, D.J. (2010). The protein tyrosine phosphatases PTPRZ and PTPRG bind to distinct members of the contactin family of neural recognition molecules. *Proc. Natl. Acad. Sci. U. S. A.* *107*, 2443–2448.
- Brankatschk, M., and Dickson, B.J. (2006). Netrins guide *Drosophila* commissural axons at short range. *Nat. Neurosci.* *9*, 188–194.
- Brose, K., Bland, K.S., Wang, K.H., Arnott, D., Henzel, W., Goodman, C.S., Tessier-Lavigne, M., and Kidd, T. (1999). Slit proteins bind Robo receptors and have an evolutionarily conserved role in repulsive axon guidance. *Cell* *96*, 795–806.
- Brumshtein, B., Greenblatt, H.M., Futerman, A.H., Silman, I., and Sussman, J.L. (2008). Control of the rate of evaporation in protein crystallization by the “microbatch under oil” method. *J. Appl. Crystallogr.* *41*, 969–971.
- Buchan, D.W.A., Minneci, F., Nugent, T.C.O., Bryson, K., and Jones, D.T. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res.* *41*, 349–357.
- Burgess, S.A., Walker, M.L., Thirumurugan, K., Trinick, J., and Knight, P.J. (2004). Use of negative stain and single-particle image processing to explore dynamic properties of flexible macromolecules. *J. Struct. Biol.* *147*, 247–258.
- Burnette, D.T., Ji, L., Schaefer, A.W., Medeiros, N.A., Danuser, G., and Forscher, P. (2008). Myosin II Activity Facilitates Microtubule Bundling in the Neuronal Growth Cone Neck. *Dev. Cell* *15*, 163–169.
- Butters, T.D., Sparks, L.M., Harlos, K., Ikemizu, S., Stuart, D.I., Jones, E.Y., and Davis, S.J. (1999). Effects of N-butyldeoxynojirimycin and the Lec3.2.8.1 mutant phenotype on N-glycan processing in Chinese hamster ovary cells: application to glycoprotein crystallization. *Protein Sci* *8*, 1696–1701.
- Carter, C.W., and Carter, C.W. (1979). Protein Crystallization Using Incomplete Factorial-Experiments. *J. Biol. Chem.* *254*, 2219–2223.
- von Castelmur, E., Marino, M., Svergun, D.I., Kreplak, L., Ucurum-Fotiadis, Z., Konarev, P. V, Urzhumtsev, A., Labeit, D., Labeit, S., and Mayans, O. (2008). A regular pattern of Ig supermotifs defines segmental flexibility as the elastic mechanism of the titin chain. *Proc. Natl. Acad. Sci. U. S. A.* *105*, 1186–1191.
- de Castro, F., Hu, L., Drabkin, H., Sotelo, C., and Chédotal, A. (1999). Chemoattraction and chemorepulsion of olfactory bulb axons by different secreted semaphorins. *J. Neurosci.* *19*, 4428–4436.
- Chan, S., Su, M.-W., Culotti, J., and Hedgecock E. (1995). unc-40 Encodes a Protein Related to the Human DCC (Deleted Colorectal Cancer) Protein. *Worm Breeder’s Gaz.* *13*. 1–5
- Chang, V.T., Crispin, M., Aricescu, A.R., Harvey, D.J., Nettleship, J.E., Fennelly, J.A., Yu, C., Boles, K.S., Evans, E.J., Stuart, D.I., Dwek, R.A., Jones, E.Y., Owens, R.J., and Davis, S.J. (2007). Glycoprotein Structural Genomics: Solving the Glycosylation Problem. *Structure* *15*, 267–273.
- Chao, D.L., Ma, L., and Shen, K. (2009). Transient cell-cell interactions in neural circuit



formation. *Nat. Rev. Neurosci.* *10*, 262–271.

Chauvet, S., and Rougon, G. (2009). The growth cone: an integrator of unique cues into refined axon guidance. *Biol. Reports* *4*, 1–4.

Chayen, N.E. (1997). A novel technique to control the rate of vapour diffusion, giving larger protein crystals. *J. Appl. Crystallogr.* *30*, 198–202.

Chayen, N., and Saridakis, E. (2008). Protein crystallization: from purified protein to diffraction-quality crystal. *Nat. Methods* *5*, 147–153.

Chayen, N.E., Shaw Stewart, P.D., Maeder, D.L., Blow, D.M. (1990). An automated system for micro-batch protein crystallization and screening. *J. Appl. Crystallogr.* *23*, 297–302.

Chédotal, A., and Richards, L.J. (2010). Wiring the brain: the biology of neuronal guidance. *Cold Spring Harb. Perspect. Biol.* *2*, 1–17.

Chen, B.E., Kondo, M., Garnier, A., Watson, F.L., Püettmann-Holgado, R., Lamar, D.R., and Schmucker, D. (2006). The molecular diversity of Dscam is functionally required for neuronal wiring specificity in *Drosophila*. *Cell* *125*, 607–620.

Chen, Q., Sun, X., Zhou, X., Liu, J., Wu, J., Zhang, Y., and Wang, J. (2013). N-terminal horseshoe conformation of DCC is functionally required for axon guidance and might be shared by other neural receptors. *J. Cell Sci.* *126*, 186–195.

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* *66*, 12–21.

Cheng, Y., Grigorieff, N., Penczek, P.A., and Walz, T. (2015). A Primer to Single-Particle Cryo-Electron Microscopy. *Cell* *161*, 438–449.

Chizhikov, V.V., and Millen, K.J. (2004). Mechanisms of roof plate formation in the vertebrate CNS. *Nat. Rev. Neurosci.* *5*, 808–812.

Cirulli, V., and Yebra, M. (2007). Netrins: beyond the brain. *Nat. Rev. Mol. Cell Biol.* *8*, 296–306.

Colamarino, S. and Tessier-Lavigne, M. (1995). The axonal chemoattractant netrin-1 is also a chemorepellent for trochlear motor axons. *Cell* *81*, 621–629.

Colavita, A. and Culotti, J.G. (1998). Suppressors of ectopic UNC-5 growth cone steering identify eight genes involved in axon guidance in *Caenorhabditis elegans*. *Dev. Biol.* *194*, 72–85.

Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. Sect. D Biol. Crystallogr.* *62*, 1002–1011.

Cowtan, K. (2008). Fitting molecular fragments into electron density. *Acta Crystallogr. Sect. D Biol. Crystallogr.* *64*, 83–89.

Cowtan, K. (2010). Recent developments in classical density modification. *Acta Crystallogr. Sect. D Biol. Crystallogr.* *66*, 470–478.

Purves, D., Augustine, J.G., Fitzpatrick, D., Katz, L.C., LaMantia, A-S., McNamara, J., and Williams, S.M. (2001). *Neuroscience* (Sinauer Associates, Inc.).

Dalva, M.B., McClelland, A.C., and Kayser, M.S. (2007). Cell adhesion molecules: signalling functions at the synapse. *Nat. Rev. Neurosci.* *8*, 206–220.

Danev, R., and Baumeister, W. (2016). Cryo-EM single-particle analysis with the volta phase plate. *Elife* *5*, 1–14.

- Davis, S.J., Puklavec, M.J., Ashford, D.A., Harlos, K., Jones, E.Y., Stuart, D.I., and Williams, A.F. (1993). Expression of soluble recombinant glycoproteins with predefined glycosylation: application to the crystallization of the T-cell glycoprotein CD2. *Prot. Engin.* *6*, 229–232.
- Deller, M.C., Kong, L., and Rupp, B. (2016). Protein stability: A crystallographer's perspective. *Acta Crystallogr. Sect. Struct. Biol. Commun.* *72*, 72–95.
- Dent, E.W., Gupton, S.L., and Gertler, F.B. (2011). The growth cone cytoskeleton in axon outgrowth and guidance. *Cold Spring Harb. Perspect. Biol.* *3*, 1–39.
- Dickson, B.J. (2002). Molecular mechanisms of axon guidance. *Science* *298*, 1959–1964.
- Duff, M.R., Grubbs, J., Howell, E.E., Duff, Jr., M.R., Grubbs, J., and Howell, E.E. (2011). Isothermal Titration Calorimetry for Measuring Macromolecule-Ligand Affinity. *J. Vis. Exp.* *55*, 5–8.
- Edelman, G.M. (1984). Cell adhesion and morphogenesis: the regulator hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* *81*, 1460–1464.
- Eglitis, M.A., and Mezey, E. (1997). Hematopoietic cells differentiate into both microglia and macroglia in the brains of adult mice. *Proc. Natl. Acad. Sci. U. S. A.* *94*, 4080–4085.
- Emsley, P., and Cowtan, K. (2004). Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* *60*, 2126–2132.
- Eroglu, C., and Barres, B.A. (2010). Regulation of synaptic connectivity by glia. *Nature* *468*, 223–231.
- Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.-Y., Pieper, U., and Sali, A. (2007). Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci. Chapter 2*, Unit 2.9.
- Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr. D. Biol. Crystallogr.* *62*, 72–82.
- Fearon, E.R., Cho, K.R., Nigro, J.M., Kern, S.E., Simons, J.W., Ruppert, J.M., Hamilton, S.R., Preisinger, A.C., Thomas, G., and Kinzler, K.W. (1990). Identification of a chromosome 18q gene that is altered in colorectal cancers. *Science* *247*, 49–56.
- Finci, L.I., Krüger, N., Sun, X., Zhang, J., Chegkazi, M., Wu, Y., Schenk, G., Mertens, H.D.T., Svergun, D.I., Zhang, Y., Wang, J.-H., and Meijers, R. (2014). The Crystal Structure of Netrin-1 in Complex with DCC Reveals the Bifunctionality of Netrin-1 As a Guidance Cue. *Neuron* *83*, 839–849.
- Finn, R.D., Coghill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J., and Bateman, A. (2016). The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* *44*, D279–D285.
- Flores, S.C., Lu, L.J., Yang, J., Carriero, N., and Gerstein, M.B. (2007). Hinge Atlas: relating protein sequence to sites of structural flexibility. *BMC Bioinformatics* *8*, 1–20.
- Flores, S.C., Keating, K.S., Painter, J., Morcos, F., Nguyen, K., Merritt, E.A., Kuhn, L.A., and Gerstein, M.B. (2008). HingeMaster: Normal mode hinge prediction approach and integration of complementary predictors. *Proteins Struct. Funct. Genet.* *73*, 299–319.
- Follmer, C., Pereira, F. V., Da Silveira, N.P., and Carlini, C.R. (2004). Jack bean urease (EC 3.5.1.5) aggregation monitored by dynamic and static light scattering. *Biophys. Chem.* *111*, 79–87.
- Fowler, S.B., and Clarke, J. (2001). Mapping the Folding Pathway of an Immunoglobulin

- Domain. *Structure* 9, 355–366.
- Frank, J. (2006). *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State* (Oxford University Press).
- Franke, D., and Svergun, D.I. (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* 42, 342–346.
- Freeze, H.H., and Kranz, C. (2010). Endoglycosidase and glycoamidase release of N-linked glycans. *Curr Protoc Mol Biol.* 17, 1–33.
- French, B.Y.S., and Wilson, K. (1978). On the Treatment of Negative Intensity Observations. *Acta Crystallogr. Sect. A Found. Crystallogr.* 34, 517–525.
- Freyer, M.W., and Lewis, E.A. (2008). Isothermal Titration Calorimetry: Experimental Design, Data Analysis, and Probing Macromolecule/Ligand Binding and Kinetic Interactions. *Methods Cell Biol.* 84, 79–113.
- Fuerst, P.G., and Burgess, R.W. (2009). Adhesion molecules in establishing retinal circuitry. *Curr. Opin. Neurobiol.* 19, 389–394.
- Fuerst, P.G., Koizumi, A., Masland, R.H., and Burgess, R.W. (2008). Neurite arborization and mosaic spacing in the mouse retina require DSCAM. *Nature* 451, 470–474.
- Gabadinho, J., Beteva, A., Guijarro, M., Rey-Bakaikoa, V., Spruce, D., Bowler, M.W., Brockhauser, S., Flot, D., Gordon, E.J., Hall, D.R., et al. (2010). MxCuBE: A synchrotron beamline control environment customized for macromolecular crystallography experiments. *J. Synchrotron Radiat.* 17, 700–707.
- Garrett, A.M., Tadenev, A.L.D., and Burgess, R.W. (2012). DSCAMs: restoring balance to developmental forces. *Front. Mol. Neurosci.* 5, 1–7.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExpASY Server. *The Proteomics Protocols Handbook* (Humana Press).
- Geisbrecht, B.V., Dowd, K.A., Barfield, R.W., Longo, P.A., and Leahy, D.J. (2003). Netrin binds discrete subdomains of DCC and UNC5 and mediates interactions between DCC and heparin. *J. Biol. Chem.* 278, 32561–32568.
- Guinier, A. (1939). La diffraction des rayons X aux tres petits angles; application a l'etude de phenomenes ultramicroscopiques. *Ann Phys.* 12, 161–237
- Gupta, R., and Brunak, S. (2002). Prediction of glycosylation across the human proteome and the correlation to protein function. *Pacific Symp. Biocomput.* 7, 310–322.
- Harrelson, A.L., and Goodman, C.S. (1988). Growth cone guidance in insects: fasciclin II is a member of the immunoglobulin superfamily. *Science* 242, 700–708.
- Harris, R., Sabatelli, L.M., and Seeger, M.A. (1996). Guidance cues at the Drosophila CNS midline: identification and characterization of two Drosophila Netrin/UNC-6 homologs. *Neuron* 17, 217–228.
- Hayward, S., and Berendsen, H.J.C. (1998). Systematic analysis of domain motions in Proteins from conformational change: new results on citrate synthase and T4 lysozyme. *Proteins Struct. Funct. Gen.* 30, 144–154.
- He, H., Kise, Y., Izadifar, A., Urwyler, O., Ayaz, D., Yan, B., Erfurth, M., Dascenco, D., and Schmucker, D. (2014). Cell-intrinsic requirement of Dscam1 isoform diversity for axon collateral formation. *Science* 344, 1182–1186.
- Hedgecock, E.M., Culotti, J.G., and Hall, D.H. (1990). The unc-5, unc-6, and unc-40 genes

guide circumferential migrations of pioneer axons and mesodermal cells on the epidermis in *C. elegans*. *Neuron* 4, 61–85.

Hoffmann, E. De., and Stroobant, V. (2007). *Mass spectrometry: principles and applications* (J. Wiley).

Hong, K., Hinck, L., Nishiyama, M., Poo, M.M., Tessier-Lavigne, M., and Stein, E. (1999). A ligand-gated association between cytoplasmic domains of UNC5 and DCC family receptors converts netrin-induced growth cone attraction to repulsion. *Cell* 97, 927–941.

Hong, K., Nishiyama, M., Henley, J., Tessier-Lavigne, M., and Poo, M.-M. (2000). Calcium signalling in the guidance of nerve growth by netrin-1. *Nature* 403, 93–98.

Huang H., Shao Q., Qu, C., Yang T., Dwyer, T., and Liu, G. (2016). Coordinated Interaction of Down Syndrome Cell Adhesion Molecule (DSCAM) and Deleted in Colorectal Cancer (DCC) with Dynamic TUBB3 Mediates Netrin-1-Induced Axon Branching. *Neuroscience* 293, 109–122.

Ishii, N., Wadsworth, W.G., Stern, B.D., Culotti, J.G., and Hedgecock, E.M. (1992). UNC-6, a laminin-related protein, guides cell and pioneer axon migrations in *C. elegans*. *Neuron* 9, 873–881.

Jancarik, J., and Kim, S.H. (1991). Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Crystallogr.* 24, 409–411.

Janin, J. (1997). Specific versus non-specific contacts in protein crystals. *Nature* 4, 973–974.

Janin, J., and Rodier, F. (1995). Protein-protein interaction at crystal contacts. *Proteins Struct. Funct. Genet.* 23, 580–587.

Jawhari, A., Uhring, M., De Carlo, S., Crucifix, C., Tocchini-Valentini, G., Moras, D., Schultz, P., and Poterszman, A. (2006). Structure and oligomeric state of human transcription factor TFIIIE. *EMBO Rep.* 7, 500–505.

Jessen, K.R. (2004). Glial cells. *Int. J. Biochem. Cell Biol.* 10, 1861–1867.

Kabsch, W. (1976). A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. Sect. A Found. Crystallogr.* 32, 922–923.

Kabsch, W. (2010). XDS. *Acta Crystallogr D Biol Crystallogr* 66, 125–132.

Kadison, S.R., Murakami, F., Matise, M.P., and Kaprielian, Z. (2006). The role of floor plate contact in the elaboration of contralateral commissural projections within the embryonic mouse spinal cord. *Dev. Biol.* 296, 499–513.

Keino-Masu, K., Masu, M., Hinck, L., Leonardo, E.D.D., Chan, S.S.-Y.Y., Culotti, J.G., and Tessier-Lavigne, M. (1996). Deleted in Colorectal Cancer (DCC) encodes a netrin receptor. *Cell* 87, 175–185.

Keleman, K., and Dickson, B.J. (2001). Short- and long-range repulsion by the *Drosophila* Unc5 Netrin receptor. *Neuron* 32, 605–617.

Kelly, L.A., Mezulis, S., Yates, C., Wass, M., and Sternberg, M. (2015). The Phyre2 web portal for protein modelling, prediction, and analysis. *Nat. Protoc.* 10, 845–858.

Kennedy, T.E., Serafini, T., de la Torre, J.R., and Tessier-Lavigne, M. (1994). Netrins are diffusible chemotropic factors for commissural axons in the embryonic spinal cord. *Cell* 78, 425–435.

Kettenmann, H., and Verkhratsky, A. (2008). Neuroglia: the 150 years after. *Trends Neurosci.* 31, 653–659.

- Kidd, T., Bland, K.S., and Goodman, C.S. (1999). Slit is the midline repellent for the robo receptor in *Drosophila*. *Cell* *96*, 785–794.
- Kim, D.E., Chivian, D., and Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* *32*, 526–531.
- Kise, Y., and Schmucker, D. (2013). Role of self-avoidance in neuronal wiring. *Curr. Opin. Neurobiol.* *23*, 983–989.
- Klämbt, C., Jacobs, J.R., and Goodman, C.S. (1991). The midline of the *Drosophila* central nervous system as a model for the genetic analysis of cell fate, cell migration and growth cone guidance. *Cell* *64*, 801–815.
- Klein, R. (2004). Eph/ephrin signaling in morphogenesis, neural development and plasticity. *Curr. Opin. Cell Biol.* *16*, 580–589.
- Ko, S.Y., Dass, C.R., and Nurgali, K. (2012). Netrin-1 in the developing enteric nervous system and colorectal cancer. *Trends Mol. Med.* *18*, 544–554.
- Kolodkin, A.L., and Tessier-Lavigne, M. (2011). Mechanisms and molecules of neuronal wiring: a primer. *Cold Spring Harb. Perspect. Biol.* *3*, 1–14.
- Kolodkin, A.L., Matthes, D.J., and Goodman, C.S. (1993). The semaphorin genes encode a family of transmembrane and secreted growth cone guidance molecules. *Cell* *75*, 1389–1399.
- Kolodziej, P.A., Timpe, L.C., Mitchell, K.J., Fried, S.R., Goodman, C.S., Jan, L.Y., and Jan, Y.N. (1996). frazzled Encodes a *Drosophila* member of the DCC immunoglobulin subfamily and is required for CNS and motor axon guidance. *Cell* *87*, 197–204.
- Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, H.J., and Svergun, D.I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Cryst.* *36*, 1277–1282.
- Kozin, M.B., and Svergun, D.I. (2001). Automated matching of high- and low-resolution structural models research papers. *J. Appl. Crystallogr.* *34*, 33–41.
- Krimpenfort, P., Song, J.-Y., Proost, N., Zevenhoven, J., Jonkers, J., and Berns, A. (2012). Deleted in colorectal carcinoma suppresses metastasis in p53-deficient mammary tumours. *Nature* *482*, 538–541.
- Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* *372*, 774–797.
- Kruger, R.P., Lee, J., Li, W., and Guan, K.-L. (2004). Mapping netrin receptor binding reveals domains of Unc5 regulating its tyrosine phosphorylation. *J. Neurosci.* *24*, 10826–10834.
- Kuroda, D., and Gray, J.J. (2016). Shape complementarity and hydrogen bond preferences in protein-protein interfaces: Implications for antibody modeling and protein-protein docking. *Bioinformatics* *32*, 2451–2456.
- Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* *227*, 680–685.
- Lai Wing Sun, K., Correia, J.P., and Kennedy, T.E. (2011). Netrins: versatile extracellular cues with diverse functions. *Development* *138*, 2153–2169.
- Langley, J.N. (1895). Note on Regeneration of Prae-Ganglionic Fibres of the Sympathetic. *J. Physiol.* *18*, 280–284.
- Lawrence, M.C., and Colman, P.M. (1993). Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* *234*, 946–950.

- Leslie, A.G.W., Powell, H.R., Read, R.J., and Sussman, J.L. (2007). Evolving Methods for Macromolecular Crystallography. *NATO Science Series* 245, 41–51.
- Leung-Hagesteijn, C., Spence, A.M., Stern, B.D., Zhou, Y., Su, M.W., Hedgecock, E.M., and Culotti, J.G. (1992). UNC-5, a transmembrane protein with immunoglobulin and thrombospondin type 1 domains, guides cell and pioneer axon migrations in *C. elegans*. *Cell* 71, 289–299.
- Levitan, I.B., and Kaczmarek, L.K. (2002). *The Neuron* (Oxford University Press).
- Li, S., Cheng, L., Yu, Y., and Chen, Q. (2016). Structural basis of Dscam1 homodimerization: Insights into context constraint for protein recognition. *Sci. Adv.* 2, 1–9.
- Lin, D.M., and Goodman, C.S. (1994). Ectopic and increased expression of Fasciclin II alters motoneuron growth cone guidance. *Neuron* 13, 507–523.
- Liu, G., Li, W., Wang, L., Kar, A., Guan, K.-L., Rao, Y., and Wu, J.Y. (2009). DSCAM functions as a netrin receptor in commissural axon pathfinding. *Proc. Natl. Acad. Sci. U. S. A.* 106, 2951–2956.
- Liu, Y., Stein, E., Oliver, T., Li, Y., Brunken, W.J., Koch, M., Tessier-Lavigne, M., and Hogan, B.L.M. (2004). Novel role for Netrins in regulating epithelial behavior during lung branching morphogenesis. *Curr. Biol.* 14, 897–905.
- Long, F., Vagin, A.A., Young, P., and Murshudov, G.N. (2008). BALBES: A molecular-replacement pipeline. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 64, 125–132.
- Lowery, L.A., and van Vactor, D. (2009). The trip of the tip: understanding the growth cone machinery. *Nat. Rev. Mol. Cell Biol.* 10, 332–343.
- Ly, A., Nikolaev, A., Suresh, G., Zheng, Y., Tessier-Lavigne, M., and Stein, E. (2008). DSCAM is a netrin receptor that collaborates with DCC in mediating turning responses to netrin-1. *Cell* 133, 1241–1254.
- Ma, Y., Acosta, D.M., Whitney, J.R., Podgornik, R., Steinmetz, N.F., French, R.H., and Parsegian, V.A. (2015). Determination of the second virial coefficient of bovine serum albumin under varying pH and ionic strength by composition-gradient multi-angle static light scattering. *J. Biol. Phys.* 41, 85–97.
- Matthews, B.J., and Grueber, W.B. (2011). Dscam1-mediated self-avoidance counters netrin-dependent targeting of dendrites in *Drosophila*. *Curr. Biol.* 21, 1480–1487.
- Matthews, B.J., Kim, M.E., Flanagan, J.J., Hattori, D., Clemens, J.C., Zipursky, S.L., and Grueber, W.B. (2007). Dendrite self-avoidance is controlled by Dscam. *Cell* 129, 593–604.
- Mazelin, L., Bernet, A., Bonod-Bidaud, C., Pays, L., Arnaud, S., Gespach, C., Bredesen, D.E., Scoazec, J.-Y., and Mehlen, P. (2004). Netrin-1 controls colorectal tumorigenesis by regulating apoptosis. *Nature* 431, 80–84.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658–674.
- McPherson, A. (2004). Introduction to protein crystallization. *Methods* 34, 254–265.
- Medeiros, N.A., Burnette, D.T., and Forscher, P. (2006). Myosin II functions in actin-bundle turnover in neuronal growth cones. *Nat. Cell Biol.* 8, 215–226.
- Meijers, R., Puettmann-Holgado, R., Skinnotis, G., Liu, J., Walz, T., Wang, J., and Schmucker, D. (2007). Structural basis of Dscam isoform specificity. *Nature* 449, 487–491.
- Meyer, R.L. (1998). Roger Sperry and his chemoaffinity hypothesis. *Neuropsychologia* 36, 957–980.

- Mörtl, M., Sonderegger, P., Diederichs, K., and Welte, W. (2007). The crystal structure of the ligand-binding module of human TAG-1 suggests a new mode of homophilic interaction. *Protein Sci.* *16*, 2174–2183.
- Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D Biol. Crystallogr.* *53*, 240–255.
- Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* *67*, 355–367.
- Nettleship, J.E., Walter, T.S., Aplin, R., Stammers, D.K., and Owens, R.J. (2005). Sample preparation and mass-spectrometric characterization of crystal-derived protein samples. *Acta Crystallogr. Sect. D Biol. Crystallogr.* *61*, 643–645.
- Neves, G., Zucker, J., Daly, M., and Chess, A. (2004). Stochastic yet biased expression of multiple Dscam splice variants by individual cells. *Nat. Genet.* *36*, 240–246.
- Nimesh, S., and Chandra, R. (2008). Guanidinium-grafted polyethylenimine: An efficient transfecting agent for mammalian cells. *Eur. J. Pharm. Biopharm.* *68*, 647–655.
- Nishiyama, M., von Schimmelmann, M.J., Togashi, K., Findley, W.M., and Hong, K. (2008). Membrane potential shifts caused by diffusible guidance signals direct growth-cone turning. *Nat. Neurosci.* *11*, 762–771.
- Odenthal, J., van Eeden, F.J., Haffter, P., Ingham, P.W., and Nüsslein-Volhard, C. (2000). Two distinct cell populations in the floor plate of the zebrafish are induced by different pathways. *Dev. Biol.* *219*, 350–363.
- Ohi, M., Li, Y., Cheng, Y., and Walz, T. (2004). Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy. *Biol. Proced. Online* *6*, 23–34.
- Padilla, J.E., and Yeates, T.O. (2003). A statistic for local intensity differences: Robustness to anisotropy and pseudo-centering and utility for detecting twinning. *Acta Crystallogr. Sect. D Biol. Crystallogr.* *59*, 1124–1130.
- Pantoliano, M.W., Petrella, E.C., Kwasnoski, J.D., Lobanov, V.S., Myslik, J., Graf, E., Carver, T., Asel, E., Springer, B.A., Lane, P. and Salemme, F.R. (2001). High-Density Miniaturized Thermal Shift Assays as a General strategy for Drug Discovery. *J. Biomol. Screen.* *6*, 429–440.
- Pape, T., and Schneider, T.R. (2004). HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *J. Appl. Crystallogr.* *37*, 843–844.
- Petoukhov, M.V., Svergun, D.I., Gerstein, M., Edwards, A., Arrowsmith, C.H., Montelione, G.T., Sali, A., Glaeser, R., Earnest, T., Baumeister, W., et al. (2005). Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.* *89*, 1237–1250.
- Petoukhov, M.V., Franke, D., Shkumatov, A.V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D.T., Konarev, P.V., and Svergun, D.I. (2012). New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* *45*, 342–350.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* *25*, 1605–1612.
- Phillips, K., and de la Peña, A.H. (2011). The combined use of the thermofluor assay and thermoQ analytical software for the determination of protein stability and buffer optimization as an aid in protein crystallization. *Curr. Protoc. Mol. Biol.* Unit 10.28, 1–15.

- Porath, J., Carlsson, J., Olsson, I., and Belfrage, G. (1975). Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* *258*, 598–599.
- Porod, G. (1982). General Theory. In *Small Angle X-Ray Scattering* (Glatter O & Kratky O, eds). Acad. Press. London, UK, 17–52.
- Prinz, M., and Priller, J. (2014). Microglia and brain macrophages in the molecular age: from origin to neuropsychiatric disease. *Nat. Rev. Neurosci.* *15*, 300–312.
- Rajasekharan, S., and Kennedy, T.E. (2009). The netrin protein family. *Gen. Biol.* *10*, 1–8.
- Ranscht, B. (2000). Cadherins: molecular codes for axon guidance and synapse formation. *Int. J. Dev. Neurosci.* *18*, 643–651.
- Rayment, I. (1997). Reductive alkylation of lysine residues to alter crystallization properties of proteins. *Methods Enzymol.* *276*, 171–179.
- Rayment, I. (2002). Small-Scale Batch Crystallization of Proteins Revisited: An Underutilized Way to Grow Large Protein Crystals. *Structure* *10*, 147–151.
- Raymond, S., and Weintraub, L. (1959). Acrylamide gel as a supporting medium for zone electrophoresis. *Science* *130*, 711.
- Reeves, P.J., Callewaert, N., Contreras, R., and Khorana, H.G. (2002). Structure and function in rhodopsin: High-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible -acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line. *PNAS* *99*, 13419–13424.
- Rhodes, G. (2000). *Crystallograph Made Crystal Clear* (Academic Press).
- Ries, C.H., Cannarile, M.A., Hoves, S., Benz, J., Wartha, K., Runza, V., Rey-Giraud, F., Pradel, L.P., Feuerhake, F., Klaman, I., et al. (2014). Targeting tumor-associated macrophages with anti-CSF-1R antibody reveals a strategy for cancer therapy. *Cancer Cell* *25*, 846–859.
- Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* *42*, W320–324.
- Saha, R.P., Bahadur, R.P., Pal, A., Mandal, S., and Chakrabarti, P. (2006). ProFace: a server for the analysis of the physicochemical features of protein-protein interfaces. *BMC Struct. Biol.* *6*, 11.
- Sawaya, M.R., Wojtowicz, W.M., Andre, I., Qian, B., Wu, W., Baker, D., Eisenberg, D., and Zipursky, S.L. (2008). A double S shape provides the structural basis for the extraordinary binding specificity of Dscam isoforms. *Cell* *134*, 1007–1018.
- Schmucker, D., and Chen, B. (2009). Dscam and DSCAM: complex genes in simple animals, complex animals yet simple genes. *Genes Dev.* *23*, 147–156.
- Schmucker, D., Clemens, J.C., Shu, H., Worby, C. A, Xiao, J., Muda, M., Dixon, J.E., and Zipursky, S.L.L. (2000). *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* *101*, 671–684.
- Serafini, T., Colamarino, S.A, Leonardo, E.D., Wang, H., Beddington, R., Skarnes, W.C., and Tessier-Lavigne, M. (1996). Netrin-1 is required for commissural axon guidance in the developing vertebrate nervous system. *Cell* *87*, 1001–1014.
- Shamsuddin, R., Doktorova, M., Jaswal, S., Lee-St John, A., and McMenimen, K. (2013). Computational prediction of hinge axes in proteins. 3rd Int. Conf. Comput. Adv. Bio Med. Sci. ICCABS *15*, 1–11.
- Sheldrick, G.M. (2008). A short history of SHELX. *Acta Crystallogr. Sect. A Found. Crystallogr.*



64, 112–122.

Shirasaki, R., Katsumata, R., and Murakami, F. (1998). Change in chemoattractant responsiveness of developing axons at an intermediate target. *Science* *279*, 105–107.

Shukla, A.K., Westfield, G.H., Xiao, K., Reis, R.I., Huang, L.-Y., Tripathi-Shukla, P., Qian, J., Li, S., Blanc, A., Oleskie, A.N., et al. (2014). Visualization of arrestin recruitment by a G-protein-coupled receptor. *Nature* *512*, 218–222.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* *539*, 1-11.

Song, Y., Dimaio, F., Wang, R.Y.R., Kim, D., Miles, C., Brunette, T., Thompson, J., and Baker, D. (2013). High-resolution comparative modeling with RosettaCM. *Structure* *21*, 1735–1742.

Sperry, R.W. (1963). Chemoaffinity in the Orderly Growth of Nerve Fiber Patterns and Connections. *Proc. Natl. Acad. Sci. U. S. A.* *50*, 703–710.

Sugimoto, Y., Taniguchi, M., Yagi, T., Akagi, Y., Nojyo, Y., and Tamamaki, N. (2001). Guidance of glial precursor cell migration by secreted cues in the developing optic nerve. *Development* *128*, 3321–3330.

Svergun, D.I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* *25*, 495–503.

Svergun, D.I., and Koch, M.H.J. (2003). Small-angle scattering studies of biological macromolecules in solution. *Rep. Prog. Phys.* *66*, 1735–1782.

Tamagnone, L., Artigiani, S., Chen, H., He, Z., Ming, G.I., Song, H., Chedotal, A., Winberg, M.L., Goodman, C.S., Poo, M., et al. (1999). Plexins are a large family of receptors for transmembrane, secreted, and GPI-anchored semaphorins in vertebrates. *Cell* *99*, 71–80.

Tessier-Lavigne, M. (1995). Eph receptor tyrosine kinases, axon repulsion, and the development of topographic maps. *Cell* *82*, 345–348.

The UniProt Consortium (2017). UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* *45*, 158–169.

Thompson, R.F., Walker, M., Siebert, C.A., Muench, S.P., and Ranson, N.A. (2016). An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology. *Methods* *100*, 3–15.

Tomaselli, K.J., Neugebauer, K.M., Bixby, J.L., Lilien, J., and Reichardt, L.F. (1988). N-cadherin and integrins: two receptor systems that mediate neuronal process outgrowth on astrocyte surfaces. *Neuron* *1*, 33–43.

Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. U. S. A.* *76*, 4350–4354.

Tracz, E., Dickson, D.W., Hainfeld, J.F., and Ksiezak-Reding, H. (1997). Paired helical filaments in corticobasal degeneration: The fine fibrillary structure with NanoVan. *Brain Res.* *773*, 33–44.

Vagin, A., and Teplyakov, A. (1997). MOLREP: an Automated Program for Molecular Replacement. *J. Appl. Crystallogr.* *30*, 1022–1025.

Vaguine, A.A., Richelle, J., and Wodak, S.J. (1999). SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr. D. Biol. Crystallogr.* *55*, 191–205.

- Vangone, A., Spinelli, R., Scarano, V., Cavallo, L., and Oliva, R. (2011). COCOMAPS: A web application to analyze and visualize contacts at the interface of biomolecular complexes. *Bioinformatics* 27, 2915–2916.
- Verstraete, K., Vandriessche, G., Januar, M., Elegheert, J., Shkumatov, A. V., Desfosses, A., Van Craenenbroeck, K., Svergun, D.I., Gutsche, I., Vergauwen, B., et al. (2011). Structural insights into the extracellular assembly of the hematopoietic Flt3 signaling complex. *Blood* 118, 60–68.
- Volkov, V.V., and Svergun, D.I. (2003). Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Cryst.* 36, 860–864.
- Walter, T.S., Meier, C., Assenberg, R., Au, K.-F., Ren, J., Verma, A., Nettleship, J.E., Owens, R.J., Stuart, D.I., and Grimes, J.M. (2006). Lysine methylation as a routine rescue strategy for protein crystallization. *Structure* 14, 1617–1622.
- Watson, F.L., Püttmann-Holgado, R., Thomas, F., Lamar, D.L., Hughes, M., Kondo, M., Rebel, V.I., and Schmucker, D. (2005). Extensive diversity of Ig-superfamily proteins in the immune system of insects. *Science* 309, 1874–1878.
- Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 67, 235–242.
- 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.
- Wu, W., Ahlsen, G., Baker, D., Shapiro, L., and Zipursky, L.S. (2012). Complementary Chimeric Isoforms Reveal Dscam1 Binding Specificity In Vivo. *Neuron* 74, 261–268.
- Xu, K., Wu, Z., Renier, N., Antipenko, A., Tzvetkova-Robev, D., Xu, Y., Minchenko, M., Nardi-Dei, V., Rajashankar, K.R., Himanen, J., et al. (2014). Structures of netrin-1 bound to two receptors provide insight into its axon guidance mechanism. *Science* 344, 1275–1279.
- Yamagata, M., and Sanes, J.R. (2008). Dscam and Sidekick proteins direct lamina-specific synaptic connections in vertebrate retina. *Nature* 451, 465–469.
- Yamagishi, S., Yamada, K., Sawada, M., Nakano, S., Mori, N., Sawamoto, K., and Sato, K. (2015). Netrin-5 is highly expressed in neurogenic regions of the adult brain. *Front. Cell. Neurosci.* 9, 1–9.
- Yamakawa, K., Huot, Y.K., Haendelt, M.A., Hubert, R., Chen, X.N., Lyons, G.E., and Korenberg, J.R. (1998). DSCAM: a novel member of the immunoglobulin superfamily maps in a Down syndrome region and is involved in the development of the nervous system. *Hum. Mol. Genet.* 7, 227–237.
- Yang, Z., Fang, J., Chittuluru, J., Asturias, F.J., and Penczek, P.A. (2012). Iterative stable alignment and clustering of 2D transmission electron microscope images. *Structure* 20, 237–247.
- Yebra, M., Montgomery, A.M.P., Diaferia, G.R., Kaido, T., Silletti, S., Perez, B., Just, M.L., Hildbrand, S., Hurford, R., Florkiewicz, E., et al. (2003). Recognition of the neural chemoattractant Netrin-1 by integrins alpha6beta4 and alpha3beta1 regulates epithelial cell adhesion and migration. *Dev. Cell* 5, 695–707.

Yuzawa, S., Opatowsky, Y., Zhang, Z., Mandiyan, V., Lax, I., and Schlessinger, J. (2007). Structural Basis for Activation of the Receptor Tyrosine Kinase KIT by Stem Cell Factor. *Cell* 130, 323–334.

Zhang, J., and Cai, H. (2010). Netrin-1 prevents ischemia/reperfusion-induced myocardial infarction via a DCC/ERK1/2/eNOS s1177/NO/DCC feed-forward mechanism. *J. Mol. Cell. Cardiol.* 48, 1060–1070.

Zinn, K. (2007). Dscam and neuronal uniqueness. *Cell* 129, 455–456.

Zipursky, S.L., and Grueber, W.B. (2013). The molecular basis of self-avoidance. *Annu. Rev. Neurosci.* 36, 547–568.



## APPENDIX

Table S1. List of laboratory consumables and equipment in alphabetical order.

Consumables & Equipment	Application	Supplier
ÄKTA FPLC, Purifier & Explorer	Liquid chromatography	GE Healthcare Life Sciences
Amicon Ultra centrifugal filter units	Protein concentration	Merck Millipore
Bio-Rad Econo-Column (150 ml bed volume)	Protein purification	Bio-Rad
Mini Trans-Blot Cell	Western blot	Bio-Rad
Chromatography columns for ÄKTA use	Protein purification	GE Healthcare Life Sciences
Corning Costar Spin-X Plastic Centrifuge Tube Filters	Filters 0.22 $\mu\text{m}$	Sigma Aldrich
Cryo-loops with easy-snap microtube	Crystallography	Hampton Research
Crystallization linbro plates (24-well)	Crystallography	Hampton Research
Crystallization plates IQ (96-well)	Crystallography	TTPLabtech
Crystallization plates Greiner (96-well)	Crystallography	TTPLabtech
Culture test plates (6-well)	Mini-scale for HEK cells	Techno Plastic Products
Culture flasks (Greiner-T175)	Midi-scale for HEK cells	Greiner Bio-One
Deep 96-well block plates	Protein elution	Thermo Fisher Scientific
Desalting Columns (Pd-10)	Desalting & buffer exchange	GE Healthcare Life Sciences
Dialysis Cassettes Slide-A-Lyzer	Buffer exchange	Thermo Fisher Scientific
Filter papers	HEK cell harvest	Carl Roth
GelDoc	Western blot	Bio-Rad
iCycler iQ Real-Time PCR detection system	Thermofluor	Bio-Rad
Malvern's OmniSEC GPC/SEC multidetector system	FPLC-SLS	Malvern Ltd, UK
Mosquito-LCP	Crystallization Robot for screen plate preparation and drop configuration	TTPLabtech
Nanodrop Spectrophotometer ND1000	Absorbance measurements for proteins and nucleic acids	NanoDrop Technologies, Wilmington, DE
Nitrocellulose Membrane Protran	Western blot	Sigma Aldrich
Paper Wicks	Crystallization	Hampton Research
Polystyrene Roller Bottles 2125 cm <sup>2</sup>	Large-scale cultures for HEK cells (250ml)	Greiner Bio-One
Mini PROTEAN Tetra Cell	SDS-PAGE	Bio-Rad
Siliconized glass cover slides	Crystallography	Hampton Research
Scorpion	Crystallization Robot for screen plate builder	ARI
Voyager DE-STR MALDI-TOF	Mass spectrometry	Applied Biosystems
Wheaton incubator (Model I057606-C)	HEK cell cultures	Wheaton Industries

Table S2. List of chemicals in alphabetical order.

<b>Chemicals</b>	<b>Application</b>	<b>Supplier</b>
Additive Screen	Crystallography	Hampton Research
Agarose	Gel electrophoresis	Serva Electrophoresis
DNA Loading Dye 6x	Loading dye	Thermo Fisher Scientific
Dulbecco's modified eagle's medium	HEK cell transfection	Merck Millipore
ECL Western blotting Substrate (Pierce™)	Western blot	Thermo Fisher Scientific
Fetal Calf Serum-Gibco (low endotoxin level)	Protein expression-HEK cells	Thermo Fisher Scientific
GeneRuler 1kb Plus	DNA Ladder	Thermo Fisher Scientific
Glutaraldehyde solution grade I, (8 % in H <sub>2</sub> O) suitable for EM	Protein cross-linking	Sigma-Aldrich
Grid Screen Salt HT	Crystallography	Hampton Research
InstantBlue coomassie stain	Single-step protein stain	Expedeon
JCSG Core Suite I,II,III,IV,plus	Crystallography	Qiagen
L-Glutamine	HEK cell transfection	Biochrom
Nickel Sepharose Excel	IMAC purification	GE Healthcare Life Sciences
Non-essential amino acids 1x	HEK cell transfection	Biochrom
NuPage LDS Sample Buffer 4x	Protein loading dye	Thermo Fisher Scientific
PageRuler Plus Prestained Protein Ladder	Protein Ladder	Thermo Fisher Scientific
Paraffin oil	Crystallography	Hampton Research
PBS Dulbecco's	HEK cell transfection	Biochrom
PCT: Pre-Crystallization Test	Crystallography	Hampton Research
PEG (powder/solution)	Crystallography	Sigma Aldrich
PGA LM_HT-96	Crystallography	Molecular Dimensions
pHClear suite I,II	Crystallography	Qiagen
Polyethylenimine (PEI)	HEK cell transfection	Sigma Aldrich
Protein Complex Suite	Crystallography	Qiagen
QIAquick Gel Extraction kit	DNA extraction from agarose gel	Qiagen
QIAprep Spin Miniprep kit	DNA extraction from bacterial cells	Qiagen
QIAplasmid Plus Gigaprep kit	DNA extraction from bacterial cells	Qiagen
RotiMark Protein Ladder	His-tagged protein ladder for Western blot	Carl Roth
Silicon oil	Crystallography	Hampton Research
SuperSignal West Pico Chemiluminescent	Western blot	Thermo Fisher Scientific
SYPRO orange dye	Thermofluor	Invitrogen
Qiagen PEG I, II suite	Crystallography	Qiagen
Talon Metal Affinity Resin	IMAC purification	Clontech Laboratories
Trypan Blue	HEK cell viability test	Biochrom
Trypsin 10x-EDTA (1:250)	Cell dissociation	Biochrom

All chemicals used for buffer preparation were of analytical grade and were purchased by Carl Roth.

Table S3. List of antibodies in alphabetical order.

Antibodies	Application	Supplier
Anti-Bovine Serum Albumin (from mouse)	Primary monoclonal antibody for Western blot	Thermo Fisher Scientific
Anti-mouse HRP conjugate (from goat)	Secondary polyclonal antibody for Western blot	Thermo Fisher Scientific
Penta-His (from mouse) BSA free	Primary monoclonal antibody for Western blot	Qiagen

Table S4. List of enzymes in alphabetical order.

Enzymes	Sequence	Application	Supplier
DpnI	5'-GA <sup>^</sup> TC-3' 3'-CT <sup>^</sup> AG-5'	Double DNA digestion	New England Biolabs
Endoglycosidase H		Deglycosylation	New England Biolabs
HindIII	5'-A <sup>^</sup> AGCTT-3' 3'-TTCGA <sup>^</sup> A-5'	Double DNA digestion	New England Biolabs
KpnI	5'-GGTAC <sup>^</sup> C-3' 3'-C <sup>^</sup> CATGG-5'	Double DNA digestion	New England Biolabs
NotI	5'-GC <sup>^</sup> GGCCGC-3' 3'-CGCCGG <sup>^</sup> CG-5'	Double DNA digestion	New England Biolabs
SacI	5'-GAGCT <sup>^</sup> C-3' 3'-C <sup>^</sup> TCGAG-5'	Double DNA digestion	New England Biolabs
Phusion High Fidelity Polymerase	-	PCR	New England Biolabs
T4 DNA Ligase	-	DNA Ligation	New England Biolabs

Table S5. List of primers for construct design.

Primer	Target	Cloning Site	Sequence
Overlap-Forward	DSCAM	Ig1-Ig9	5'-ACGTGTCCAAAAGCATGTACCTGACTGTGAAGATCCCCGCTATGATTACTTCCTATCCCAA-3'
Overlap-Reverse	DSCAM	Ig1-Ig9	5'-AGTAATCATAGCGGGGATCTTCACAGTCAGGTACATGCTTTTGGACACGTCGGCCC-3'
D9-Reverse	DSCAM	Ig1-Ig9	5'-GTGGTGGTGGAGCTCTTCCTGGACGGTCAGCTGGA-3'
Forward	DSCAM	Ig1-Ig4	5'-TAGGTACCCATAGTTCCTGTATTTTGTGAACGCC-3'
Reverse	DSCAM	Ig1-Ig4	5'-ATAGAGCTCCTCCAGGACCACCTGCACGTAGT-3'

Table S6. SAXS data collection and scattering parameters for DSCAM Ig1-Ig8 WT.

<b>Data collection parameters</b>	
Beamline	P12 - PETRA III (EMBL-Hamburg DESY, Germany)
Beam geometry (mm <sup>2</sup> )	0.2 x 0.12
Wavelength (Å)	1.24
s-range (Å <sup>-1</sup> )	0.008 - 0.475
Exposure time (ms)	50
Concentration range (mg ml <sup>-1</sup> )	1.5 - 7.5
Temperature (K)	283
<b>Structural parameters</b>	
<i>I</i> (0) (arbitrary units) [from p(r)]	5926±50
<i>R</i> <sub>g</sub> (Å) [from p(r)]	5.9±0.2
<i>I</i> (0) (arbitrary units) (from Guinier)	5894.1±45
<i>R</i> <sub>g</sub> (Å) (from Guinier)	5.8±0.2
<i>D</i> <sub>max</sub> (Å)	20±0.5
Porod volume estimate (Å <sup>3</sup> )	494±50
Dry volume calculated from sequence (Å <sup>3</sup> )	399
<b>Molecular mass determination (Da)</b>	
Absolute calibration	167000±17000
Porod volume	309000±30000
BSA	200000±20000
Theoretical monomeric based on sequence	88183

Table S7. SAXS data collection and scattering parameters for DSCAM Ig1-Ig9 WT.

<b>Data collection parameters</b>	
Beamline	P12 - PETRA III (EMBL-Hamburg DESY, Germany)
Beam geometry (mm <sup>2</sup> )	0.2 x 0.12
Wavelength (Å)	1.24
s-range (Å <sup>-1</sup> )	0.008 - 0.457
Exposure time (ms)	50
Concentration range (mg ml <sup>-1</sup> )	0.46 - 8.30
Temperature (K)	283
<b>Structural parameters</b>	
<i>I</i> (0) (arbitrary units) [from p(r)]	5714±50
<i>R</i> <sub>g</sub> (Å) [from p(r)]	6.4±0.2
<i>I</i> (0) (arbitrary units) (from Guinier)	1680±20
<i>R</i> <sub>g</sub> (Å) (from Guinier)	6.4±0.2
<i>D</i> <sub>max</sub> (Å)	22.3±0.5
Porod volume estimate (Å <sup>3</sup> )	726±70
Dry volume calculated from sequence (Å <sup>3</sup> )	934
<b>Molecular mass determination (Da)</b>	
Absolute calibration	226000±20000
Porod volume	454000±45000
BSA	207000±20000
Theoretical monomeric based on sequence	96890



Table S8. SAXS data collection and scattering parameters for DSCAM Ig1-Ig9:netrin<sub>VIV</sub>.

<b>Data collection parameters</b>	
Beamline	P12 - PETRA III (EMBL-Hamburg DESY, Germany)
Beam geometry (mm <sup>2</sup> )	0.2 x 0.12
Wavelength (Å)	1.24
s-range (Å <sup>-1</sup> )	0.008 - 0.457
Exposure time (ms)	50
Concentration range (mg ml <sup>-1</sup> )	0.6 - 4.0
Temperature (K)	283
<b>Structural parameters</b>	
<i>I</i> (0) (arbitrary units) [from p(r)]	4547±40
<i>R</i> <sub>g</sub> (Å) [from p(r)]	6.30±0.1
<i>I</i> (0) (arbitrary units) (from Guinier)	1300±10
<i>R</i> <sub>g</sub> (Å) (from Guinier)	6.3±0.2
<i>D</i> <sub>max</sub> (Å)	20.5±0.5
Porod volume estimate (Å <sup>3</sup> )	474±45
Dry volume calculated from sequence (Å <sup>3</sup> )	314
<b>Molecular mass determination (Da)</b>	
Absolute calibration	178000±15000
Porod volume	296000±30000
BSA	163000±15000
Theoretical monomeric based on sequence for DSCAM Ig1-Ig9:netrin <sub>VIV</sub>	146179

Table S9. SAXS data collection and scattering parameters for DSCAM Ig1-Ig4 WT.

<b>Data collection parameters</b>	
Beamline	P12 - PETRA III (EMBL-Hamburg DESY, Germany)
Beam geometry (mm <sup>2</sup> )	0.2 x 0.12
Wavelength (Å)	1.24
s-range (Å <sup>-1</sup> )	0.008 - 0.475
Exposure time (ms)	50
Concentration range (mg ml <sup>-1</sup> )	0.5 - 10.1
Temperature (K)	283
<b>Structural parameters</b>	
<i>I</i> (0) (arbitrary units) [from p(r)]	1889±16
<i>R</i> <sub>g</sub> (Å) [from p(r)]	2.9±0.1
<i>I</i> (0) (arbitrary units) (from Guinier)	1884.3±15
<i>R</i> <sub>g</sub> (Å) (from Guinier)	2.9±0.1
<i>D</i> <sub>max</sub> (Å)	10±0.1
Porod volume estimate (Å <sup>3</sup> )	71±5
Dry volume calculated from sequence (Å <sup>3</sup> )	101
<b>Molecular mass determination (Da)</b>	
Absolute calibration	53000±5000
Porod volume	44000±4000
BSA	51000±5000
Theoretical monomeric based on sequence	43754

Table S10. List of amino acids.

<b>Amino acids</b>	<b>Name</b>
Ala (A)	Alanine
Cys (C)	Cysteine
Asp (D)	Aspartate
Glu (E)	Glutamate
Phe (F)	Phenylalanine
Gly (G)	Glycine
His (H)	Histidine
Ile (I)	Isoleucine
Lys (K)	Lysine
Leu (L)	Leucine
Met (M)	Methionine
Asn (N)	Asparagine
Pro (P)	Proline
Gln (Q)	Glutamine
Arg (R)	Arginine
Ser (S)	Serine
Thr (T)	Threonine
Val (V)	Valine
Try (W)	Tryptophan
Tyr (Y)	Tyrosine

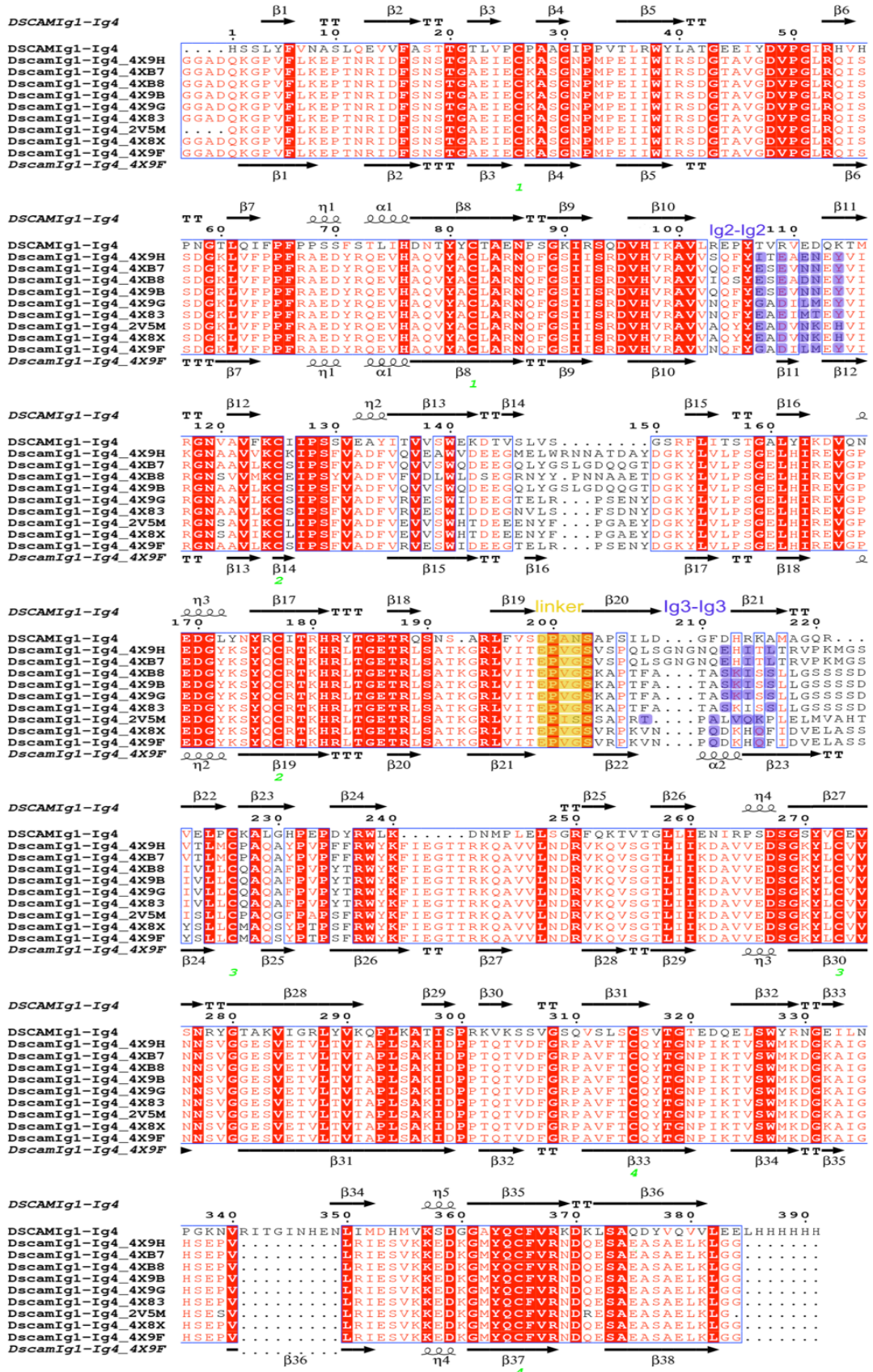


Figure S1. Sequence alignment of DSCAM Ig1-Ig4 from *H. sapiens* with Dscam Ig1-Ig4 different isoforms from *D. melanogaster* (Li et al., 2016; Meijers et al., 2007). The amino acids involved in the dimer interface between Ig2.A-Ig2.B and Ig3.A-Ig3.B in Dscam in *Drosophila*, are highlighted in purple. The conserved identical amino acids are white on a red background,

similar residues are red on a white background and the disulfide bonds are indicated in green numbers. The secondary-structure elements  $\alpha$ -helices,  $3_{10}$ -helices,  $\beta$ -strands and strict  $\beta$ -turns are denoted  $\eta$ ,  $\alpha$ ,  $\beta$  and TT, respectively. Multiple sequence alignment was performed using the ClustalW2 server (Sievers et al., 2011) and the output was processed and visualized using the ESPript 3.0 server (Robert and Gouet, 2014).

## ACKNOWLEDGEMENTS

*The thesis is dedicated to Evangelia, Nikolas and my Parents.*

I would like to express my gratitude to my PhD supervisor, Dr. Rob Meijers, for giving me the opportunity to conduct my research on different challenging projects in his group and learn how to develop skills for advancing my work during the PhD. I would also like to thank him for his advice and feedback on the thesis completion and for his support all these years.

Besides my supervisor, I would also like to sincerely thank my thesis advisory committee members, Prof. Dr. Ilme Schlichting for her guidance and advice all these years, Dr. Thomas R. Schneider for his constructive comments, Dr. Francesca Peri for her support and especially Prof. Dr. Robert B. Russell, for being a member of my defense committee.

Dr. Dmitri Svergun's suggestions and guidance have been most valuable for my research project during my PhD and contributed significantly to improve my understanding on SAXS. I am also grateful to Dr. Haydyn Mertens for the SAXS model generation and the fruitful discussions and to Dr. Anne Tuukkanen, my very good friend and colleague, for her constant support, encouragement and scientific advice on SAXS. I would also like to thank Dr. Cy Jeffries for his scientific guidance during FPLC-RALS/RI/UV experiments.

I would also like to acknowledge Prof. Dr. George Skiniotis from Life Sciences Institute-University of Michigan, for his assistance and the stimulating discussions on the EM aspects of the project, as well as his group members Jeffrey Tarrasch and Annie Dosey who collected the EM data and performed the single-particle analysis. Prof. Dr. Thomas Walz, Department of Cell Biology, Harvard Medical School, has also been a great help during preliminary EM data collection.

I am extremely thankful to Dr. Isabel Bento for her contribution in the structural studies of DSCAM both during crystallographic data collection and structure determination.

A big thanks goes to all Meijers Group members and good friends that supported my research. In particular, the research technicians Dr. Nina Krüger and Alex Kötter, MSc who passed their knowledge to me on how to handle mammalian cells, Dr. Matthew Dunne, Dr. Maria Garcia Alai, Dr. Anna Gieras and Dr. Heidi Kaljunen for all the moments we shared during our lab-life and outside of it. In addition, all staff

member of EMBL-Hamburg and especially the lab manager Dr. Annabel Parret, the research technicians Dana Komadina and Morlin Milewski, MSc and the SPC facility colleagues, Dr. Stephane Boivin, Janina Hinrichs, Sandra Kozak, MSc, Maria Ioana Nemtanu and Vanessa Vieira, MSc for their help and support in the lab. A big thanks goes also to Dr. Vivian Pogenberg for his advice on crystallographic experiments. I am also grateful to Dr. Gleb Bourenkov, Dr. Michele Cianci, Dr. Guillaume Pompidor and all the staff scientists at EMBL-beamlines for their support and advice any time of day or night, along with the staff members at ESRF-Grenoble beamlines.

I am forever grateful to Dr. Spyros Chatziefthimiou for his constant advice and will to share his experience on X-ray protein crystallography and his continuous support and friendship. By working next to him in the lab, I learned everything about Ig domain-related proteins, and definitely captured the spirit of old classic Greek films. There are no words to describe the help and advice I received from Dr. Johanna Kallio. Her constant and unwearied support has been most valuable not only during crystallographic data collection but also during our endless discussions.

A very special thanks goes to Dr. Helke Hillebrand, former Dean of EMBL Graduate Studies, for all the discussions we had during the PhD representative meetings and for her constant advice and moral guidance and to Dr. Monika Lachner, Dean of Graduate Studies, who has been a great help and support. I am always grateful to EMBL-Hamburg administration and especially to Margret Fischer, our Margret, who made me immediately feel part of the EMBL family as well as to the members of the Graduate Office in Heidelberg for the friendly atmosphere and their excellent support, along with the PhD Office-Faculty of Biosciences at University of Heidelberg, for all their help.

I am also thankful to my good friends Natasha, Selina, Koen, Nabil, Daniel, Claudia and Diana for their positive thoughts and for sharing all the funny moments in the lab.

Without the advice from Dr. Evangelia D. Chrysina, the encouragement and faith from Nikolas during the journey we share, the support from my closest friend and her family, Dr. Aikaterini Chajistamatiou, the funny moments with my dearest friend Dimitra, the constant advice from Chara and Julie and my mother's support, who has always been by my side, this research journey would not have been possible.

I am forever grateful for their presence in my life.

