W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Aug 8, 2016 – 04:21 PM EDT

PDB ID	:	5L5L
Title	:	Plexin A4 full extracellular region, domains 1 to 8 modeled, data to 8 angstrom, spacegroup P2(1)
Authors	:	Janssen, B.J.C.; Kong, Y.; Malinauskas, T.; Vangoor, V.R.; Coles, C.H.; Kaufmann, R.; Ni, T.; Gilbert, R.J.C.; Padilla-Parra, S.; Pasterkamp, R.J.; Jones, E.Y.
Deposited on	:	2016-05-28
Resolution	:	8.00 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

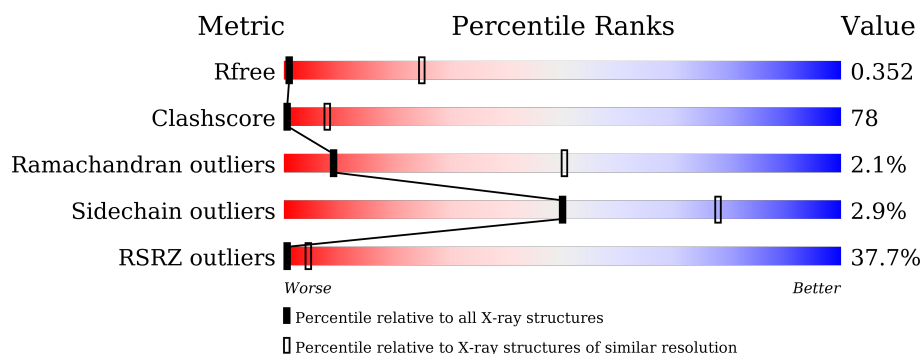|		
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027939
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027939

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 8.00 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

Metric	Percentile Ranks	Value
Rfree		0.352
Clashscore		78
Ramachandran outliers		2.1%
Sidechain outliers		2.9%
RSRZ outliers		37.7%

■ Percentile relative to all X-ray structures
▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	1015 (11.50-3.66)
Clashscore	102246	1064 (11.50-3.70)
Ramachandran outliers	100387	1036 (11.50-3.66)
Sidechain outliers	100360	1006 (11.50-3.66)
RSRZ outliers	91569	1014 (11.50-3.66)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	1207	33% 27% 52% • 17%
1	B	1207	27% 25% 47% • 24%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 15030 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Plexin-A4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1000	Total 7841	C 4938	N 1356	O 1482	S 65	0	0	0
1	B	915	Total 7189	C 4533	N 1239	O 1357	S 60	0	0	0

There are 26 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	33	GLU	-	expression tag	UNP Q80UG2
A	34	THR	-	expression tag	UNP Q80UG2
A	35	GLY	-	expression tag	UNP Q80UG2
A	1230	GLY	-	expression tag	UNP Q80UG2
A	1231	ARG	-	expression tag	UNP Q80UG2
A	1232	THR	-	expression tag	UNP Q80UG2
A	1233	LYS	-	expression tag	UNP Q80UG2
A	1234	HIS	-	expression tag	UNP Q80UG2
A	1235	HIS	-	expression tag	UNP Q80UG2
A	1236	HIS	-	expression tag	UNP Q80UG2
A	1237	HIS	-	expression tag	UNP Q80UG2
A	1238	HIS	-	expression tag	UNP Q80UG2
A	1239	HIS	-	expression tag	UNP Q80UG2
B	33	GLU	-	expression tag	UNP Q80UG2
B	34	THR	-	expression tag	UNP Q80UG2
B	35	GLY	-	expression tag	UNP Q80UG2
B	1230	GLY	-	expression tag	UNP Q80UG2
B	1231	ARG	-	expression tag	UNP Q80UG2
B	1232	THR	-	expression tag	UNP Q80UG2
B	1233	LYS	-	expression tag	UNP Q80UG2
B	1234	HIS	-	expression tag	UNP Q80UG2
B	1235	HIS	-	expression tag	UNP Q80UG2
B	1236	HIS	-	expression tag	UNP Q80UG2
B	1237	HIS	-	expression tag	UNP Q80UG2
B	1238	HIS	-	expression tag	UNP Q80UG2

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	1239	HIS	-	expression tag	UNP Q80UG2

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: Plexin-A4

• **Molecule 1: Plexin-A4**

Chain B:
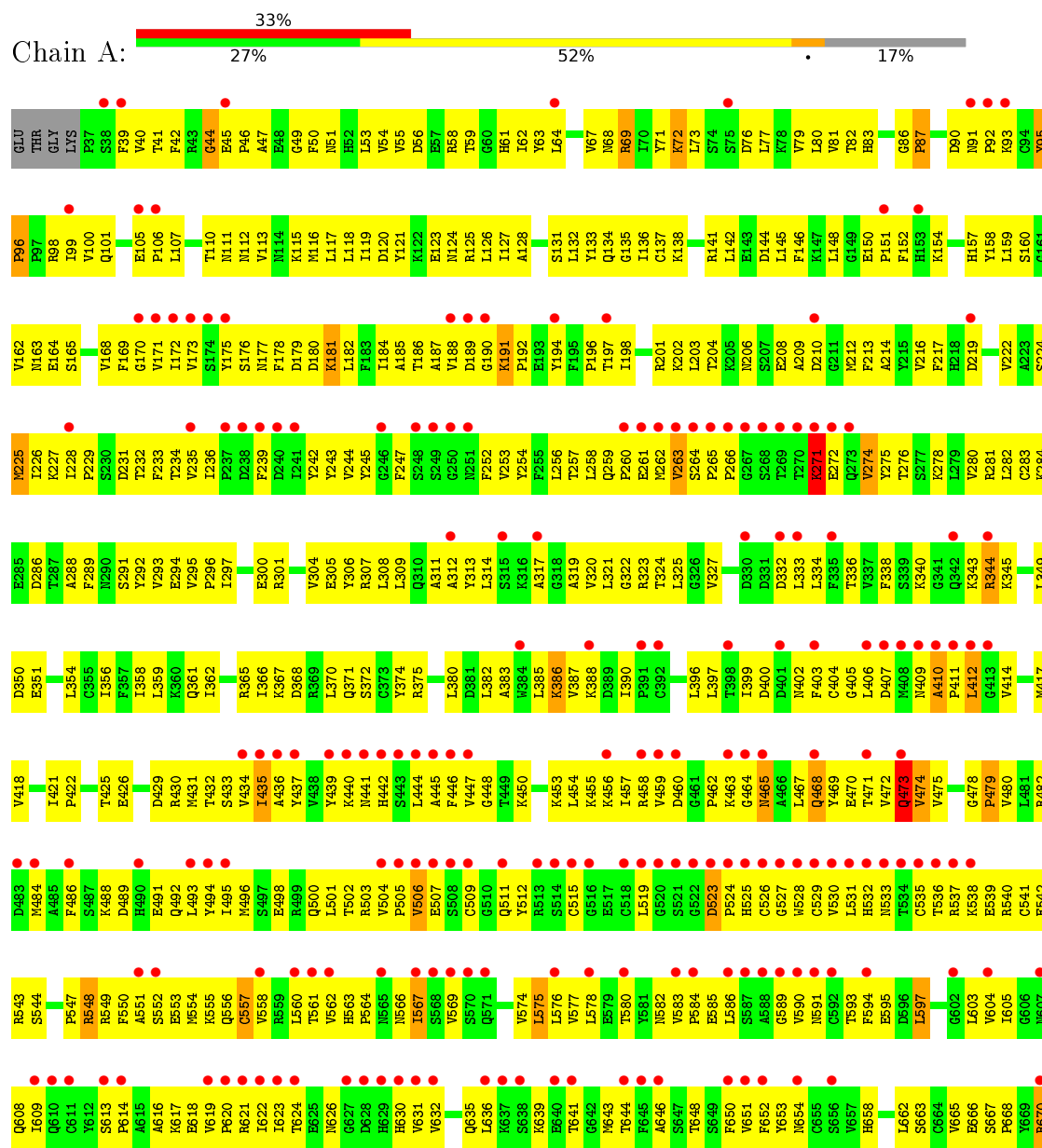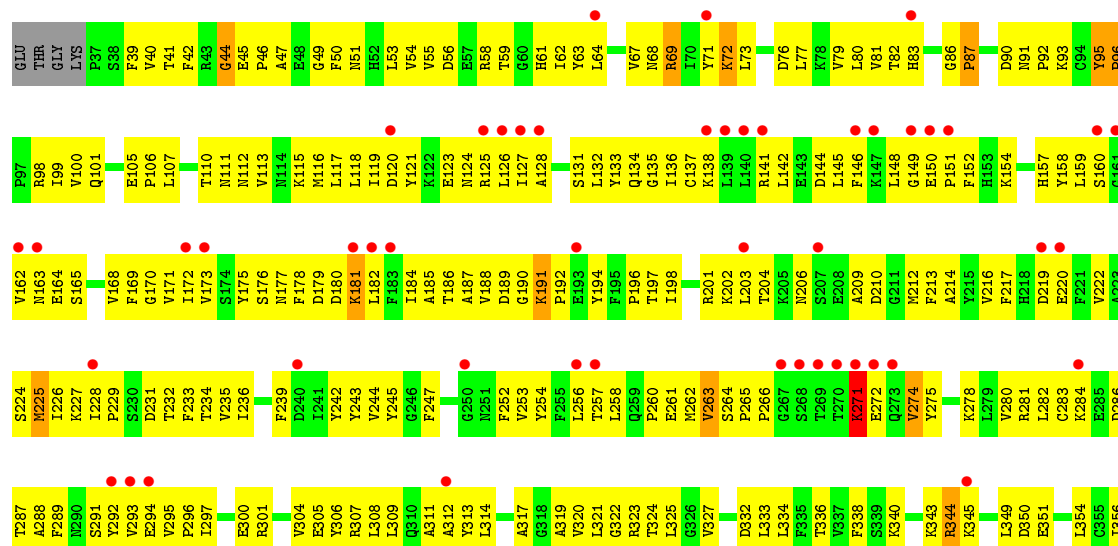
4   Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	142.18Å 241.00Å 144.07Å 90.00° 99.83° 90.00°	Depositor
Resolution (Å)	47.74 – 8.00 47.74 – 8.00	Depositor EDS
% Data completeness (in resolution range)	99.1 (47.74-8.00) 99.5 (47.74-8.00)	Depositor EDS
R_{merge}	0.16	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.18 (at 8.32Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.2_1309)	Depositor
R, R_{free}	0.349 , 0.349 0.344 , 0.352	Depositor DCC
R_{free} test set	488 reflections (4.86%)	DCC
Wilson B-factor (Å^2)	450.9	Xtriage
Anisotropy	0.543	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.40 , 548.1	EDS
L-test for twinning[2]	$< |L| > = 0.43$, $< L^2 > = 0.25$	Xtriage
Estimated twinning fraction	0.043 for l,-k,h	Xtriage
F_o,F_c correlation	0.70	EDS
Total number of atoms	15030	wwPDB-VP
Average B, all atoms (Å^2)	264.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.04% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality (i)

5.1 Standard geometry (i)

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#$|Z|$>5	RMSZ	#$|Z|$>5
1	A	1.01	5/8007 (0.1%)	1.36	27/10846 (0.2%)
1	B	1.00	5/7344 (0.1%)	1.32	24/9943 (0.2%)
All	All	1.01	10/15351 (0.1%)	1.34	51/20789 (0.2%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	3
1	B	0	4
All	All	0	7

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	506	VAL	C-N	23.13	1.87	1.34
1	B	506	VAL	C-N	21.10	1.82	1.34
1	A	557	CYS	C-N	-20.82	0.86	1.34
1	B	557	CYS	C-N	-17.13	0.94	1.34
1	A	700	CYS	C-N	-15.63	1.04	1.34

The worst 5 of 51 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	747	GLN	CG-CD-OE1	-38.83	43.94	121.60
1	A	747	GLN	CG-CD-OE1	-38.81	43.98	121.60
1	A	653	TYR	O-C-N	-33.42	69.23	122.70
1	B	557	CYS	O-C-N	-31.89	71.68	122.70
1	A	653	TYR	CA-C-N	23.37	168.61	117.20

There are no chirality outliers.

5 of 7 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	854	CYS	Mainchain
1	A	863	ILE	Peptide
1	A	95	TYR	Peptide
1	B	557	CYS	Mainchain
1	B	95	TYR	Peptide

5.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	7841	0	7710	1244	29
1	B	7189	0	7050	1075	59
All	All	15030	0	14760	2319	61

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 78.

The worst 5 of 2319 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:868:PRO:HD2	1:A:981:GLY:CA	1.32	1.52
1:A:868:PRO:CD	1:A:981:GLY:CA	1.87	1.50
1:A:873:THR:CA	1:A:982:SER:HB2	1.46	1.43
1:A:873:THR:HA	1:A:982:SER:CB	1.48	1.40
1:B:506:VAL:HG22	1:B:525:HIS:NE2	1.33	1.38

The worst 5 of 61 symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:146:PHE:CE1	1:B:730:GLN:CD[1_655]	0.64	1.56
1:B:287:THR:OG1	1:B:840:HIS:CG[1_655]	0.67	1.53
1:A:146:PHE:CE1	1:B:730:GLN:OE1[1_655]	0.77	1.43
1:A:146:PHE:CD1	1:B:730:GLN:OE1[1_655]	0.78	1.42

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:287:THR:CA	1:B:840:HIS:NE2[1_655]	0.79	1.41

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	994/1207 (82%)	923 (93%)	51 (5%)	20 (2%)	9	51
1	B	907/1207 (75%)	845 (93%)	43 (5%)	19 (2%)	9	50
All	All	1901/2414 (79%)	1768 (93%)	94 (5%)	39 (2%)	9	50

5 of 39 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	96	PRO
1	A	181	LYS
1	A	191	LYS
1	A	410	ALA
1	A	465	ASN

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	888/1067 (83%)	861 (97%)	27 (3%)	48	77
1	B	812/1067 (76%)	789 (97%)	23 (3%)	51	78

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	1700/2134 (80%)	1650 (97%)	50 (3%)	50 78

5 of 50 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	892	HIS
1	B	69	ARG
1	B	806	MET
1	A	1004	ILE
1	A	1017	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 53 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	892	HIS
1	B	157	HIS
1	B	789	ASN
1	A	970	GLN
1	A	1006	ASN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	B	6
1	A	5

The worst 5 of 11 chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	B	854:CYS	C	855:THR	N	2.49
1	A	802:LYS	C	803:CYS	N	2.46
1	A	951:MET	C	952:THR	N	2.32
1	B	653:TYR	C	654:ASN	N	2.31
1	B	802:LYS	C	803:CYS	N	2.01

6 Fit of model and data (i)

6.1 Protein, DNA and RNA chains (i)

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2			OWAB(Å²)	Q<0.9
1	A	1000/1207 (82%)	1.99	393 (39%)	0	4	195, 258, 410, 410	0
1	B	915/1207 (75%)	1.81	328 (35%)	0	4	209, 257, 329, 329	0
All	All	1915/2414 (79%)	1.90	721 (37%)	0	4	195, 257, 329, 410	0

The worst 5 of 721 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1029	GLN	16.5
1	A	269	THR	16.2
1	A	1018	VAL	14.5
1	A	270	THR	13.8
1	A	271	LYS	13.1

6.2 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates (i)

There are no carbohydrates in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.