



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 12, 2020 – 11:09 pm BST

PDB ID : 6PNP  
Title : Crystal structure of the splice insert-free neurexin-1 LNS2 domain in complex with neurexophilin-1  
Authors : Wilson, S.C.; White, K.I.; Zhou, Q.; Brunger, A.T.  
Deposited on : 2019-07-02  
Resolution : 1.94 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

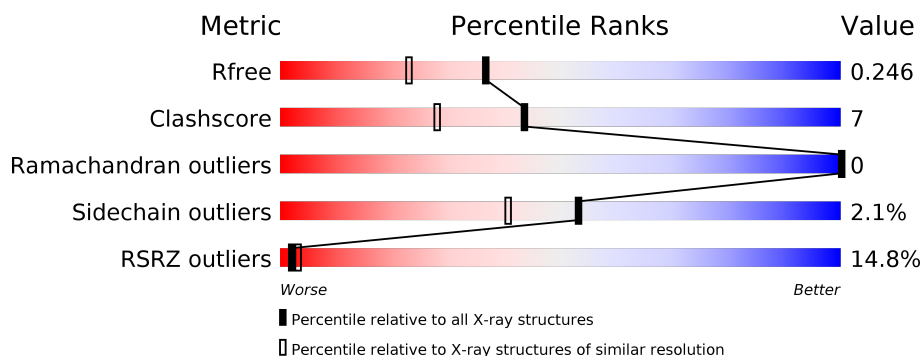
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.94 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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	4310 (1.96-1.92)
Clashscore	141614	1023 (1.94-1.94)
Ramachandran outliers	138981	1007 (1.94-1.94)
Sidechain outliers	138945	1007 (1.94-1.94)
RSRZ outliers	127900	4250 (1.96-1.92)

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  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria respectively. 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  $\leq 5\%$ . 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	194	<div> <div>14%</div> <div> <div></div> <div>88%</div> <div>12%</div> <div>.</div> </div> </div>
2	B	168	<div> <div>12%</div> <div> <div></div> <div>62%</div> <div>17%</div> <div>21%</div> </div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 5175 atoms, of which 2519 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 Neurexin-1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	193	Total	C	H	N	O	S	0	1	0
			2943	938	1460	247	292	6			

There are 27 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	280	ASP	-	expression tag	UNP Q9CS84
A	281	ALA	-	expression tag	UNP Q9CS84
A	282	SER	-	expression tag	UNP Q9CS84
A	293	ALA	CYS	engineered mutation	UNP Q9CS84
A	?	-	HIS	deletion	UNP Q9CS84
A	?	-	SER	deletion	UNP Q9CS84
A	?	-	GLY	deletion	UNP Q9CS84
A	?	-	ILE	deletion	UNP Q9CS84
A	?	-	GLY	deletion	UNP Q9CS84
A	?	-	HIS	deletion	UNP Q9CS84
A	?	-	ALA	deletion	UNP Q9CS84
A	?	-	MET	deletion	UNP Q9CS84
A	?	-	VAL	deletion	UNP Q9CS84
A	?	-	ASN	deletion	UNP Q9CS84
A	?	-	LYS	deletion	UNP Q9CS84
A	?	-	LEU	deletion	UNP Q9CS84
A	?	-	HIS	deletion	UNP Q9CS84
A	?	-	CYS	deletion	UNP Q9CS84
A	?	-	SER	deletion	UNP Q9CS84
A	481	ALA	-	expression tag	UNP Q9CS84
A	482	ALA	-	expression tag	UNP Q9CS84
A	483	LEU	-	expression tag	UNP Q9CS84
A	484	GLU	-	expression tag	UNP Q9CS84
A	485	VAL	-	expression tag	UNP Q9CS84
A	486	LEU	-	expression tag	UNP Q9CS84
A	487	PHE	-	expression tag	UNP Q9CS84
A	488	GLN	-	expression tag	UNP Q9CS84

- Molecule 2 is a protein called Neurexophilin-1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	132	Total	C	H	N	O	S	0	5	0
			2165	709	1059	180	211	6			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	110	ASP	-	expression tag	UNP Q63366
B	111	ALA	-	expression tag	UNP Q63366
B	112	ALA	-	expression tag	UNP Q63366
B	113	GLN	-	expression tag	UNP Q63366
B	114	PRO	-	expression tag	UNP Q63366
B	115	ALA	-	expression tag	UNP Q63366
B	116	ALA	-	expression tag	UNP Q63366
B	117	ARG	-	expression tag	UNP Q63366
B	118	ASP	-	expression tag	UNP Q63366
B	146	ASP	ASN	engineered mutation	UNP Q63366
B	156	ASP	ASN	engineered mutation	UNP Q63366
B	162	ASP	ASN	engineered mutation	UNP Q63366
B	272	LEU	-	expression tag	UNP Q63366
B	273	GLU	-	expression tag	UNP Q63366
B	274	VAL	-	expression tag	UNP Q63366
B	275	LEU	-	expression tag	UNP Q63366
B	276	PHE	-	expression tag	UNP Q63366
B	277	GLN	-	expression tag	UNP Q63366

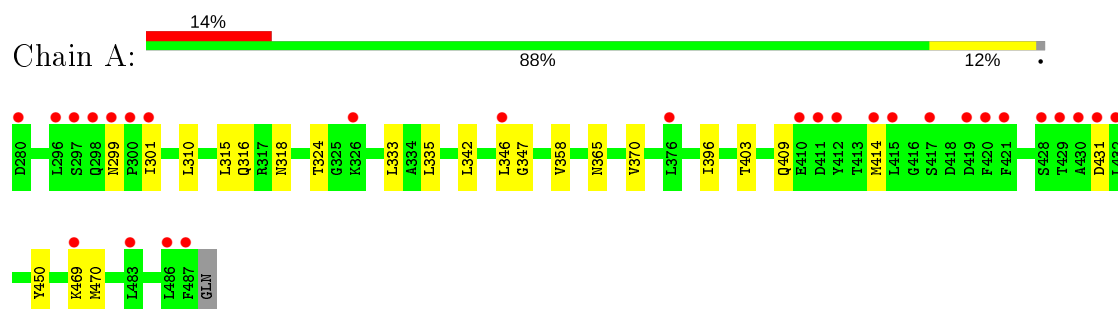
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	27	Total	O	0	0
			27	27		
3	B	40	Total	O	0	0
			40	40		

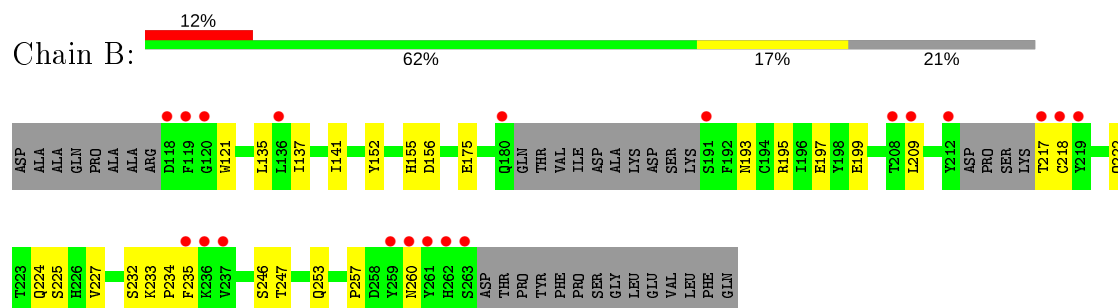
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes 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: Neurexin-1



#### • Molecule 2: Neurexophilin-1



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	71.70Å 60.98Å 79.62Å 90.00° 106.74° 90.00°	Depositor
Resolution (Å)	45.60 – 1.94 45.60 – 1.94	Depositor EDS
% Data completeness (in resolution range)	95.0 (45.60-1.94) 95.1 (45.60-1.94)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.11 (at 1.94Å)	Xtriage
Refinement program	PHENIX 1.14 _3260	Depositor
R, $R_{free}$	0.198 , 0.247 0.198 , 0.246	Depositor DCC
$R_{free}$ test set	2000 reflections (8.59%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	39.4	Xtriage
Anisotropy	0.301	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 62.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	5175	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	84.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.62% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality

### 5.1 Standard geometry

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	0.44	0/1510	0.58	0/2046
2	B	0.53	0/1133	0.62	0/1533
All	All	0.48	0/2643	0.59	0/3579

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-close contacts

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	1483	1460	1459	16	0
2	B	1106	1059	1054	22	1
3	A	27	0	0	0	0
3	B	40	0	0	2	0
All	All	2656	2519	2513	34	1

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 7.

The worst 5 of 34 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)
2:B:121:TRP:CE2	2:B:141[B]:ILE:HD11	2.18	0.78

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:121:TRP:CZ2	2:B:141[B]:ILE:HD11	2.25	0.70
2:B:197:GLU:OE1	3:B:301:HOH:O	2.13	0.65
1:A:347:GLY:O	2:B:260:ASN:ND2	2.31	0.64
1:A:358:VAL:HG21	2:B:247:THR:O	2.02	0.59

All (1) 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)
2:B:152:TYR:HH	2:B:175:GLU:OE1[4_555]	1.59	0.01

## 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	192/194 (99%)	186 (97%)	6 (3%)	0	100	100
2	B	131/168 (78%)	126 (96%)	5 (4%)	0	100	100
All	All	323/362 (89%)	312 (97%)	11 (3%)	0	100	100

There are no Ramachandran outliers to report.

### 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	164/164 (100%)	160 (98%)	4 (2%)	49	36
2	B	128/153 (84%)	126 (98%)	2 (2%)	62	52
All	All	292/317 (92%)	286 (98%)	6 (2%)	53	41

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	414	MET
2	B	222	GLN
1	A	431	ASP
1	A	409	GLN
2	B	195	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

There are no ligands in this entry.

## 5.7 Other polymers ⓘ

There are no such residues in this entry.

## 5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

## 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, 95<sup>th</sup> 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(Å <sup>2</sup> )	Q<0.9
1	A	193/194 (99%)	0.88	28 (14%) 2 3	33, 69, 161, 206	0
2	B	132/168 (78%)	0.69	20 (15%) 2 3	33, 54, 157, 198	0
All	All	325/362 (89%)	0.80	48 (14%) 2 3	33, 65, 160, 206	0

The worst 5 of 48 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	411	ASP	7.8
1	A	487	PHE	7.4
1	A	298	GLN	7.4
2	B	262	HIS	6.6
1	A	280	ASP	5.9

### 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.