



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 07:43 pm GMT

PDB ID : 5L5M  
Title : Plexin A4 full extracellular region, domains 1 to 7 modeled, data to 8 angstrom, spacegroup P4(3)2(1)2  
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)

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

<http://wwpdb.org/validation/2016/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) : 1.9-1692  
EDS : trunk28620  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
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) : recalc28949

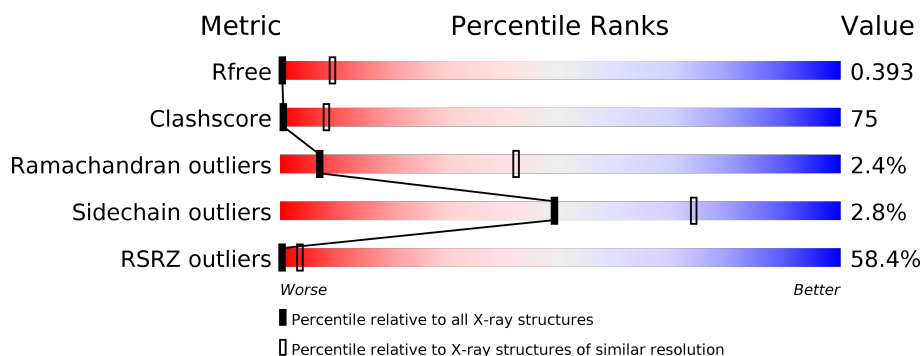
# 1 Overall quality at a glance

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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	100719	1100 (10.00-3.70)
Clashscore	112137	1036 (11.50-3.80)
Ramachandran outliers	110173	1004 (11.50-3.76)
Sidechain outliers	110143	1099 (11.50-3.70)
RSRZ outliers	101464	1004 (11.50-3.72)

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. 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	1207	<div> <div>44%</div> <div>25% 47% . . 24%</div> </div>

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 7189 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	915	Total	C	N	O	S	0	0	0
			7189	4533	1239	1357	60			

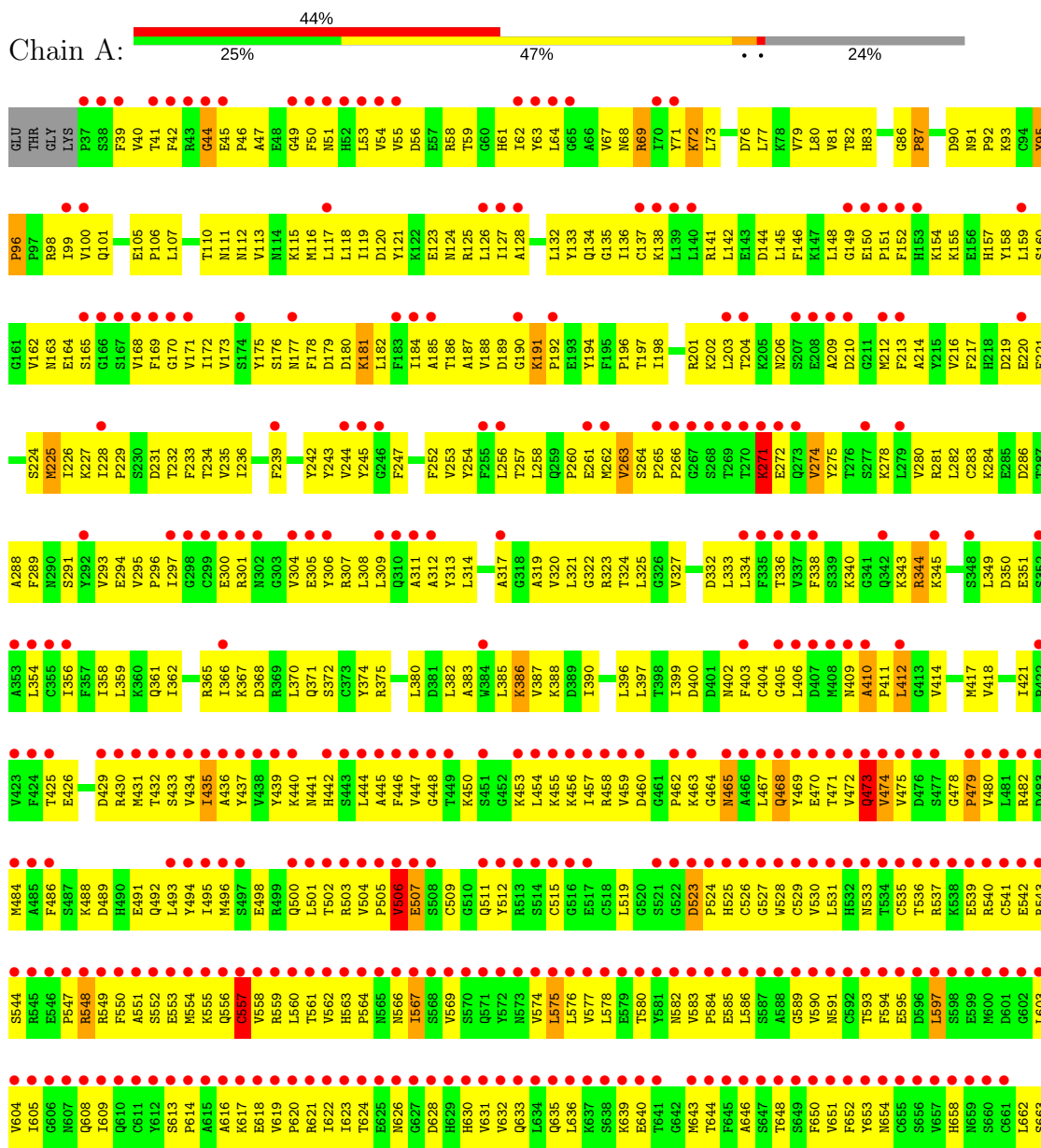
There are 13 discrepancies between the modelled and reference sequences:

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

### 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 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: Plexin-A4



GLY	PRO	GLN	VAL	THR	ASP16	N856	N792	L796	C864
GLY	GLY	ALA	GLU	ASN	H917	P857	N796	P727	V665
MET	THR	ALA	PRO	LEU	G918	R858	K797	Q728	E666
THR	ILE	ALA	THR	ALA	A920	I859	R798	P729	S667
SER	ILE	ALA	ILE	GLY	K921	E861	V799	Q730	P668
PRO	LEU	LEU	VAL	SER	P922	I862	L800	S731	Y669
GLY	LYS	GLY	ARG	ASN	G923	I863	Y801	Q732	R670
MET	GLY	PRO	ILE	VAL	Q924	P864	K802	Q733	C671
VAL	VAL	ASP	GLU	VAL	H925	V865	G803	E736	W673
TYR	ASN	HIS	PRO	VAL	A926	T866	G804	E737	C674
ILE	LEU	GLN	GLU	MET	G927	G867	A805	G738	K675
ALA	SER	SER	THR	PHE	F928	P868	M806	I739	Y676
PRO	ASP	ASP	SER	GLY	V929	R869	R807	L740	R677
GLY	PRO	LEU	ILE	SER	E930	G870	E808	W741	H678
ARG	VAL	THR	GLN	GLN	I931	G871	S809	V742	V679
THR	ALA	THR	GLU	PRO	C932	T872	C810	Q743	C680
LYS	GLY	ARG	GLY	CYS	V933	T873	G811	Q744	T681
HIS	GLY	PRO	ASN	LEU	A934	K874	L812	I745	H682
HIS	ASN	GLU	THR	PHE	V935	V875	C813	E746	D683
HIS	VAL	GLU	PRO	HIS	C936	T876	L814	Q747	P684
HIS	LEU	PHE	ILE	ARG	R937	I877	K815	R748	N685
HIS	GLY	GLY	ALA	ARG	P938	R878	A816	V749	T686
HIS	ASN	PHE	VAL	VAL	E939	G879	F750	P750	
THR	TYR	ILE	TRP	PRO	F940	E880	A751	A751	F689
THR	THR	LEU	GLY	SER	H941	N881	F820	L752	Q680
VAL	VAL	ASP	THR	TYR	A942	L882	E821	R753	E691
LEU	ASN	HIS	HIS	ILE	R943	G883	C822	F754	G692
VAL	VAL	VAL	LEU	ILE	S944	L884	G823		R693
GLY	GLN	GLN	ASP	CYS	S945	E885	W824		V694
SER	LEU	LEU	LEU	ASN	O946	F886	C825		K695
LYS	LYS	LEU	ILE	THR	L947	R887	Q826		L696
PRO	CYS	LEU	GLN	THR	Y948	C888	S827		P697
CYS	THR	ILE	ASN	SER	Y949	I889	P828		
THR	THR	LEU	PRO	SER	F950	A890	G829		C700
VAL	VAL	ASN	GLN	GLU	N951	S891	Q830		P701
THR	THR	LYS	ILE	GLU		H892	C831		Q702
VAL	VAL	THR	ARG	VAL	LEU	V893	T832		L703
ASN	ASN	ALA	ALA	LEU	THR	K894	L833		L704
ASP	ASP	PHE	LYS	LEU	LEU	A896			R705
VAL	VAL	THR	GLY	LYS	ASP	G897	H836		V706
GLN	GLN	TYR	GLY	VAL	LEU	V898	C837		D707
LEU	LEU	PRO	LYS	THR	LYS	E899	P838		K708
CYS	CYS	ASN	GLU	VAL	PRO	C900	A839		L709
GLU	GLU	PRO	HIS	GLN	ASN	S901	H840		L710
SER	SER	VAL	ILE	VAL	ARG	P902	E841		P711
PRO	PRO	PHE	ASN	VAL	GLY	L903	R843		P712
ASN	ASN	GLU	ILE	ARG	PRO	V904	W844		V713
LEU	LEU	ALA	CYS	ALA	MET	D905	L845		E714
ILE	ILE	PHE	GLU	ARG	SER	G906	E846		V715
GLY	GLY	SER	VAL	GLY	GLY	Y907	L847		I716
ARG	ARG	PRO	LEU	ARG	GLY	I908	S848		K717
HIS	HIS	PRO	GLN	GLN	THR	P909	G849		P718
LYS	LYS	GLY	ALA	ASP	GLN	A910	A850		I719
VAL	VAL	ILE	THR	LEU	VAL	E911	R851		T720
MET	MET	LEU	GLU	VAL	THR	Q912	S852		L721
ALA	ALA	GLU	MET	PHE	ILE	I913	K853		K722
ARG	ARG	LEU	THR	THR	THR	V914	C854		A723
VAL	VAL	LYS	CYS	TYR	GLY	C915	T855		K724
									N725

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	271.48Å 271.48Å 251.25Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	72.13 – 8.00 72.13 – 8.00	Depositor EDS
% Data completeness (in resolution range)	99.4 (72.13-8.00) 99.3 (72.13-8.00)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.60 (at 8.39Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.2_1309)	Depositor
R, $R_{free}$	0.373 , 0.395 0.371 , 0.393	Depositor DCC
$R_{free}$ test set	495 reflections (4.81%)	DCC
Wilson B-factor (Å <sup>2</sup> )	530.0	Xtriage
Anisotropy	0.475	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.41 , 489.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.39$ , $\langle L^2 \rangle = 0.21$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.70	EDS
Total number of atoms	7189	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	250.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.43% 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	0.97	4/7346 (0.1%)	1.46	29/9949 (0.3%)

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	7

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	557	CYS	C-N	17.01	1.73	1.34
1	A	700	CYS	C-N	-13.28	1.09	1.34
1	A	49	GLY	CA-C	6.34	1.61	1.51
1	A	49	GLY	C-N	5.06	1.45	1.34

The worst 5 of 29 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	747	GLN	CG-CD-OE1	-38.83	43.94	121.60
1	A	506	VAL	O-C-N	-35.49	65.92	122.70
1	A	557	CYS	CA-C-N	-34.27	41.81	117.20
1	A	557	CYS	C-N-CA	-32.12	41.40	121.70
1	A	854	CYS	O-C-N	-27.52	78.66	122.70

There are no chirality outliers.

5 of 7 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	506	VAL	Mainchain

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Group
1	A	700	CYS	Mainchain
1	A	802	LYS	Mainchain,Peptide
1	A	854	CYS	Mainchain
1	A	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	7189	0	7045	1066	13
All	All	7189	0	7045	1066	13

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 75.

The worst 5 of 1066 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:439:TYR:CD1	1:A:537:ARG:NH1	1.70	1.55
1:A:439:TYR:CE1	1:A:537:ARG:NH1	1.71	1.54
1:A:548:ARG:CG	1:A:584:PRO:HA	1.31	1.52
1:A:556:GLN:HA	1:A:582:ASN:CB	1.40	1.51
1:A:548:ARG:HG3	1:A:584:PRO:CA	1.45	1.40

The worst 5 of 13 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:233:PHE:O	1:A:234:THR:OG1[8_665]	1.24	0.96
1:A:83:HIS:CE1	1:A:731:SER:OG[4_455]	1.63	0.57
1:A:148:LEU:O	1:A:728:GLN:NE2[4_455]	1.65	0.55
1:A:146:PHE:CE1	1:A:730:GLN:OE1[4_455]	1.79	0.41
1:A:234:THR:CA	1:A:234:THR:CA[8_665]	1.89	0.31

## 5.3 Torsion angles

### 5.3.1 Protein backbone

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	911/1207 (76%)	843 (92%)	46 (5%)	22 (2%)	7	42

5 of 22 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

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	812/1067 (76%)	789 (97%)	23 (3%)	49	74

5 of 23 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	548	ARG
1	A	575	LEU
1	A	854	CYS
1	A	567	ILE
1	A	597	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 29 such

sidechains are listed below:

Mol	Chain	Res	Type
1	A	626	ASN
1	A	672	HIS
1	A	836	HIS
1	A	629	HIS
1	A	678	HIS

### 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	A	3

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	653:TYR	C	654:ASN	N	2.02

*Continued on next page...*

*Continued from previous page...*

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	557:CYS	C	558:VAL	N	1.73
1	A	700:CYS	C	701:PRO	N	1.09

## 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	915/1207 (75%)	4.00	534 (58%) <b>0</b> <b>3</b>	145, 231, 394, 394	0

The worst 5 of 534 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	859	ILE	29.9
1	A	860	THR	27.4
1	A	861	GLU	27.1
1	A	862	ILE	25.3
1	A	919	GLU	21.9

### 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.