wwPDB X-ray Structure Validation Summary Report (i)

May 22, 2020 – 02:01 pm BST

PDB ID	:	5L5N
Title	:	Plexin A4 full extracellular region, domains 1 to 7 modeled, data to 8.5 angstrom, spacegroup P4(3)22
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	:	8.50 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

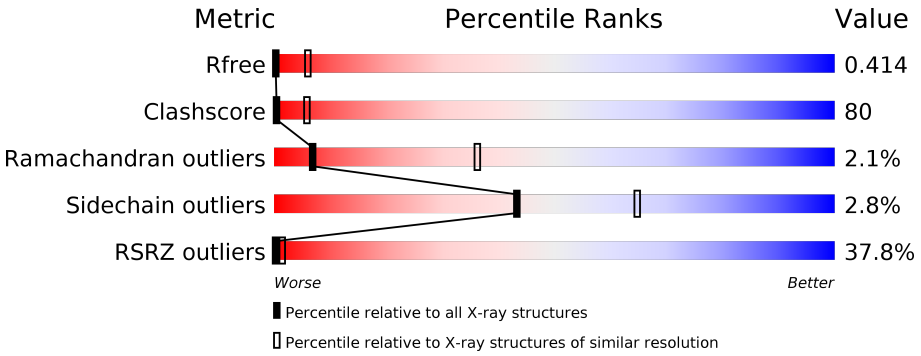
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation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

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 8.50 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	1005 (11.50-3.90)
Clashscore	141614	1070 (11.50-3.90)
Ramachandran outliers	138981	1003 (11.50-3.90)
Sidechain outliers	138945	1003 (11.50-3.86)
RSRZ outliers	127900	1004 (9.50-3.80)

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 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 <=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	29% 24% 48% . 24%

2 Entry composition (i)

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
			Total	C	N	O	S			
1	A	915	7189	4533	1239	1357	60	0	0	0

There are 13 discrepancies between the modelled and reference sequences:

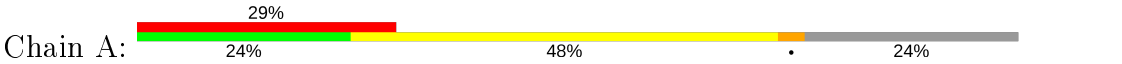
Chain	Residue	Modelled	Actual	Comment	Reference
A	33	GLU	-	expression tag	UNP Q80UG2
A	34	THR	-	expression tag	UNP Q80UG2
A	35	GLY	-	expression tag	UNP Q80UG2
A	1230	GLY	-	expression tag	UNP Q80UG2
A	1231	ARG	-	expression tag	UNP Q80UG2
A	1232	THR	-	expression tag	UNP Q80UG2
A	1233	LYS	-	expression tag	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

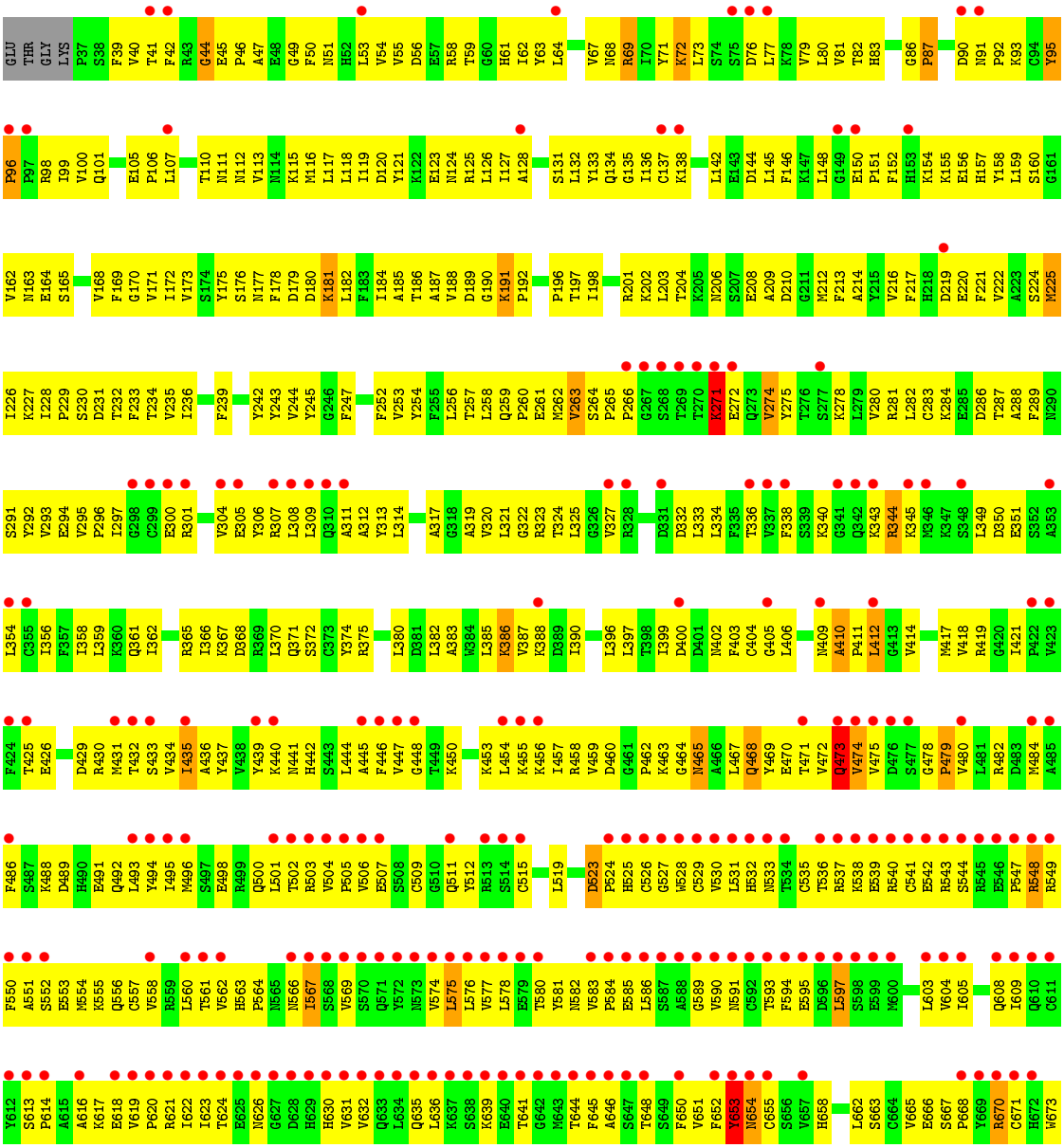
3 Residue-property plots (i)

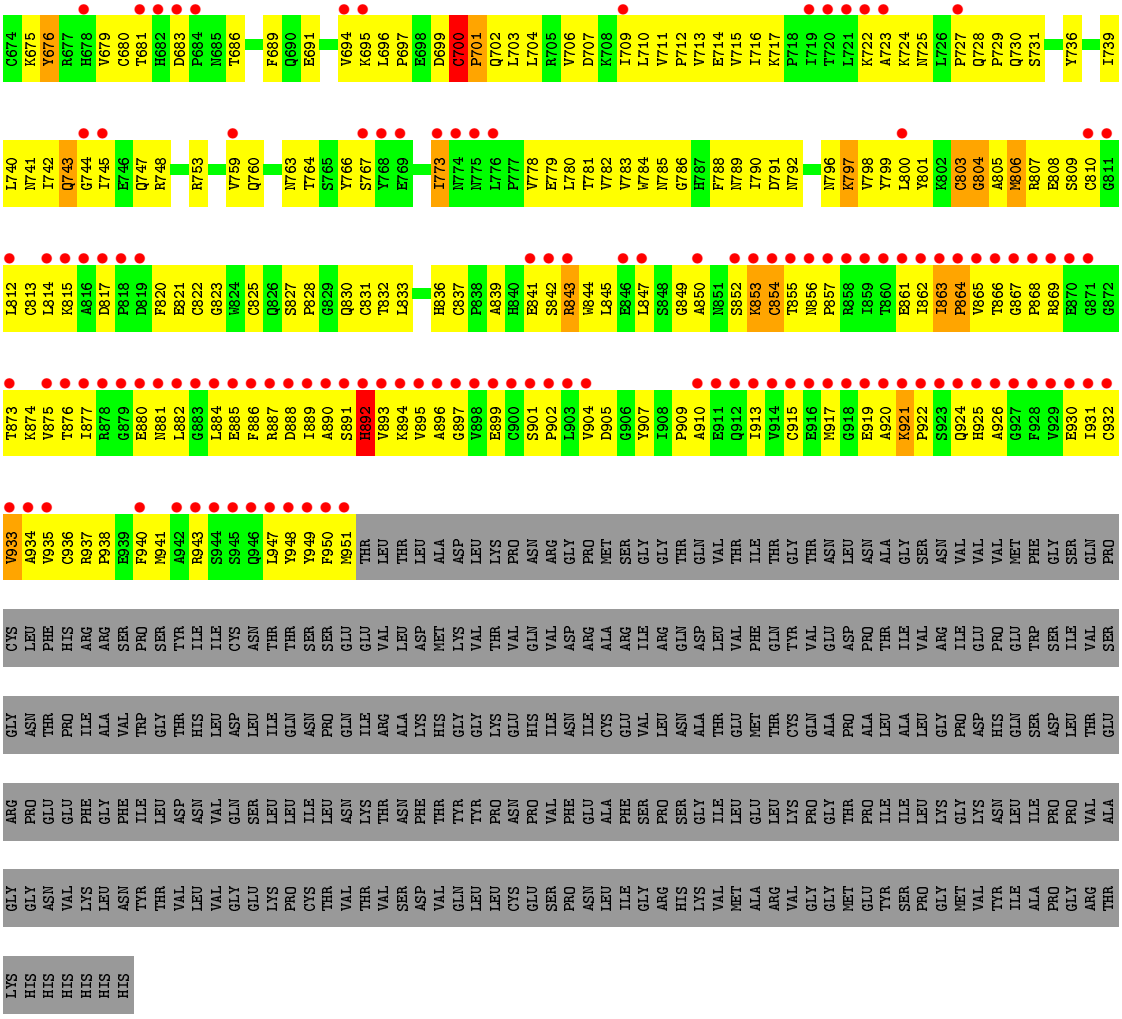
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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: Plexin-A4

Chain A:






C674 K675 Y676 R677 H678 V679 C680 T681 H682 D683 P684 N685 T686 F689 Q690 E691 V694 K695 L696 P697 E698 D699 C700 P701 Q702 L703 L704 R705 V706 D707 K708 I709 L710 V711 P712 V713 E714 V715 I716 K717 P718 I719 T720 L721 K722 A723 K724 N725 L726 P727 Q728 P729 Q730 S731 Y736 I739
L740 N741 I742 Q743 G744 I745 E746 Q747 R748 R753 V759 Q760 N763 T764 S765 Y766 S767 Y768 E769 I773 N774 N775 L776 P777 V778 E779 L780 T781 V782 V783 W784 N785 G786 H787 F788 N789 I790 D791 N792 N796 K797 V798 Y799 L800 Y801 K802 C803 G804 A805 M806 R807 E808 S809 C810 G811
L812 C813 L814 K815 A816 D817 P818 D819 F820 E821 C822 G823 W824 C825 Q826 S827 P828 G829 Q830 C831 T832 L833 H836 C837 P838 A839 H840 E841 S842 R843 W844 L845 E846 L847 S848 G849 A850 N851 S852 K853 C854 T855 N856 P857 R858 I859 T860 E861 I862 I863 P864 V865 T866 G867 P868 R869 E870 G871 G872
T873 K874 V875 T876 I877 R878 G879 E880 N881 L882 G883 L884 E885 F886 R887 D888 I889 A890 S891 H892 V893 K894 V895 A896 G897 V898 E899 C900 S901 P902 L903 V904 D905 G906 Y907 I908 P909 A910 E911 Q912 I913 V914 C915 E916 M917 G918 E919 A920 K921 P922 S923 Q924 H925 A926 G927 F928 V929 E930 I931 C932
V933 A934 V935 C936 R937 P938 E939 F940 M941 A942 R943 S944 S945 Q946 L947 Y948 Y949 F950 M951 THR LEU THR LEU ALA ASP LEU LYS PRO ASN ARG GLY PRO MET SER GLY GLY GLY THR GLN VAL THR ILE THR GLY THR ASN LEU ASN ALA GLY SER ASN VAL VAL VAL MET PHE GLY SER GLN PRO
CYS LEU PHE HIS ARG ARG SER PRO SER TYR ILE ILE CYS ASN THR THR SER SER GLU GLU VAL LEU ASP MET LYS VAL THR VAL GLN VAL ASP ARG ALA ARG ILE ARG GLN ASP LEU VAL PHE GLN TYR VAL GLU ASP PRO THR ILE VAL ARG ILE GLU GLU PRO GLU TRP SER ILE VAL THR GLU
GLY ASN THR PRO ILE ALA VAL TRP GLY THR HIS LEU ASP LEU GLN ASN PRO GLN ILE ARG ALA LYS HIS GLY GLY LYS GLU HIS ILE ASN ILE CYS GLU VAL LEU ASN ALA THR GLU MET THR CYS GLN ALA PRO ALA LEU ALA LEU LEU GLY PRO ASP HIS GLN SER ASP LEU THR GLU
ARG PRO GLU GLU PHE GLY PHE ILE LEU ASP ASN VAL GLN SER LEU LEU ILE LEU ASN LYS THR ASN PHE THR TYR TYR PRO ASN PRO VAL PHE GLU ALA PHE SER PRO SER GLY ILE LEU GLU LYS PRO GLY THR PRO ILE ILE LEU LYS GLY LYS ASN LEU ILE PRO PRO VAL ALA
GLY GLY ASN VAL LYS LEU ASN TYR THR VAL LEU VAL GLY GLU LYS PRO CYS THR VAL THR VAL SER ASP VAL GLN LEU LEU CYS GLU SER PRO ASN LEU ILE GLY ARG HIS LYS VAL MET ALA ARG VAL GLY GLY MET GLU TYR SER PRO GLY MET VAL VAL TYR ILE ALA PRO GLY ARG THR
LYS HIS HIS HIS HIS HIS HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 2 2	Depositor
Cell constants a, b, c, α, β, γ	189.48Å 189.48Å 252.59Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	70.37 – 8.50 70.37 – 8.50	Depositor EDS
% Data completeness (in resolution range)	99.5 (70.37-8.50) 99.6 (70.37-8.50)	Depositor EDS
R_{merge}	0.33	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.94 (at 8.39Å)	Xtriage
Refinement program	PHENIX 1.8.2_1309	Depositor
R, R_{free}	0.401 , 0.412 0.411 , 0.414	Depositor DCC
R_{free} test set	203 reflections (4.64%)	wwPDB-VP
Wilson B-factor (Å^2)	502.9	Xtriage
Anisotropy	0.294	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 149.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.34$, $< L^2 > = 0.17$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.53	EDS
Total number of atoms	7189	wwPDB-VP
Average B, all atoms (Å^2)	162.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.07% 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

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	1.10	4/7343 (0.1%)	1.26	20/9940 (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	4

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	653	TYR	C-N	-36.36	0.50	1.34
1	A	700	CYS	C-N	-33.68	0.70	1.34
1	A	49	GLY	CA-C	6.43	1.62	1.51
1	A	49	GLY	C-N	5.09	1.45	1.34

The worst 5 of 20 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	747	GLN	CG-CD-OE1	-38.83	43.94	121.60
1	A	653	TYR	O-C-N	-20.66	89.64	122.70
1	A	700	CYS	C-N-CD	-19.18	78.41	120.60
1	A	700	CYS	O-C-N	-10.23	101.67	121.10
1	A	747	GLN	CG-CD-NE2	-9.56	93.74	116.70

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	653	TYR	Mainchain

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
1	A	700	CYS	Mainchain
1	A	863	ILE	Peptide
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	7049	1133	32
All	All	7189	0	7049	1133	32

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 80.

The worst 5 of 1133 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:468:GLN:HG3	1:A:524:PRO:CD	1.18	1.58
1:A:530:VAL:HG11	1:A:584:PRO:CD	1.11	1.56
1:A:439:TYR:CE2	1:A:538:LYS:NZ	1.78	1.51
1:A:468:GLN:CG	1:A:524:PRO:HD3	1.39	1.51
1:A:439:TYR:HE2	1:A:538:LYS:NZ	1.11	1.44

The worst 5 of 32 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:234:THR:CA	1:A:234:THR:CA[5_455]	0.86	1.34
1:A:233:PHE:O	1:A:234:THR:OG1[5_455]	0.87	1.33
1:A:146:PHE:CE1	1:A:730:GLN:OE1[4_555]	1.03	1.17
1:A:146:PHE:CE1	1:A:730:GLN:CD[4_555]	1.26	0.94
1:A:146:PHE:CD1	1:A:730:GLN:OE1[4_555]	1.26	0.94

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	905/1207 (75%)	842 (93%)	44 (5%)	19 (2%)	7 36

5 of 19 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%)	43 65

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 27 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	163	ASN

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

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	A	6

The worst 5 of 6 chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	802:LYS	C	803:CYS	N	4.06

Continued on next page...

Continued from previous page...

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	854:CYS	C	855:THR	N	3.30
1	A	557:CYS	C	558:VAL	N	2.81
1	A	506:VAL	C	507:GLU	N	2.58
1	A	700:CYS	C	701:PRO	N	0.70

6 Fit of model and data

6.1 Protein, DNA and RNA chains

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^{th} 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^2)	Q<0.9
1	A	915/1207 (75%)	1.97	346 (37%)	0	1	100, 150, 216, 216	0

The worst 5 of 346 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	584	PRO	11.0
1	A	859	ILE	9.9
1	A	585	GLU	9.7
1	A	902	PRO	9.5
1	A	860	THR	9.3

6.2 Non-standard residues in protein, DNA, RNA chains

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

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.