



wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 15, 2020 – 11:59 PM EST

PDB ID : 7KGV
Title : Crystal structure of sodium-coupled neutral amino acid transporter SLC38A9
in the N-terminal plugged form
Authors : Lei, H.; Mu, X.; Hattne, J.; Gonen, T.
Deposited on : 2020-10-19
Resolution : 3.40 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 ⓘ 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 : 2.15.1
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) : 2.15.1

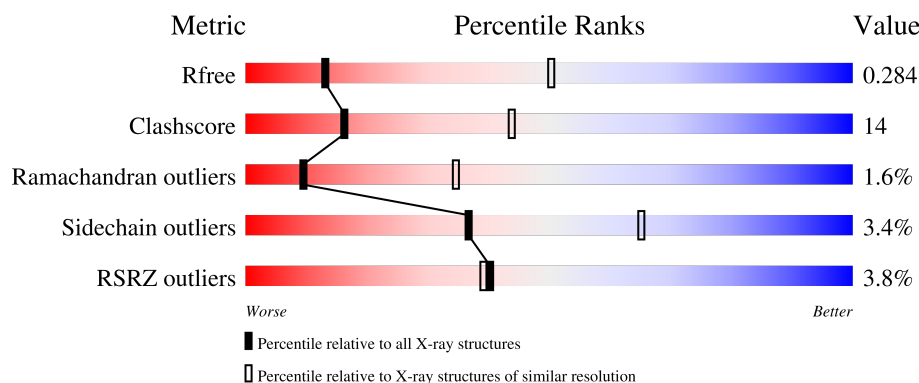
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.40 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	1026 (3.48-3.32)
Clashscore	141614	1055 (3.48-3.32)
Ramachandran outliers	138981	1038 (3.48-3.32)
Sidechain outliers	138945	1038 (3.48-3.32)
RSRZ outliers	127900	2173 (3.50-3.30)

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 of 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 $\leq 5\%$. 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	549	<div> <div>5%</div> <div>46% 17% 34%</div> </div>
1	B	549	<div> <div>3%</div> <div>47% 24% 26%</div> </div>
2	C	219	<div> <div>2%</div> <div>72% 25%</div> </div>
2	D	219	<div> <div>4%</div> <div>72% 26%</div> </div>
3	E	215	<div> <div>4%</div> <div>66% 31%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	215	 72%25%•

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 12613 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 Sodium-coupled neutral amino acid transporter 9.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	361	Total	C	N	O	S	0	0	0
			2905	1975	441	471	18			
1	B	404	Total	C	N	O	S	0	0	0
			3269	2211	508	531	19			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	227	GLN	ASN	engineered mutation	UNP Q08BA4
A	235	GLN	ASN	engineered mutation	UNP Q08BA4
A	252	GLN	ASN	engineered mutation	UNP Q08BA4
A	263	GLN	ASN	engineered mutation	UNP Q08BA4
B	227	GLN	ASN	engineered mutation	UNP Q08BA4
B	235	GLN	ASN	engineered mutation	UNP Q08BA4
B	252	GLN	ASN	engineered mutation	UNP Q08BA4
B	263	GLN	ASN	engineered mutation	UNP Q08BA4

- Molecule 2 is a protein called Monoclonal antibody Fab heavy chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	217	Total	C	N	O	S	0	0	0
			1574	987	261	319	7			
2	D	216	Total	C	N	O	S	0	0	0
			1569	984	260	318	7			

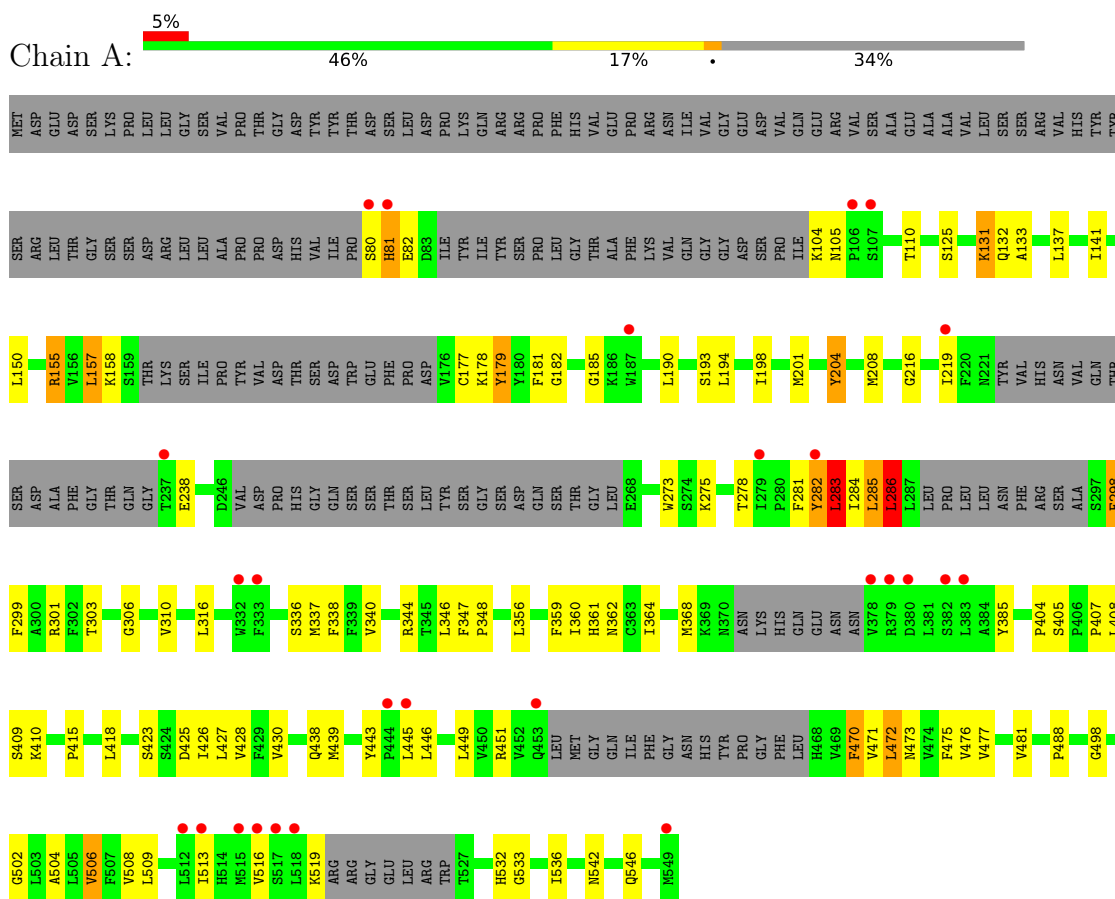
- Molecule 3 is a protein called Monoclonal antibody Fab light chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	E	215	Total	C	N	O	S	0	0	0
			1648	1036	273	335	4			
3	F	215	Total	C	N	O	S	0	0	0
			1648	1036	273	335	4			

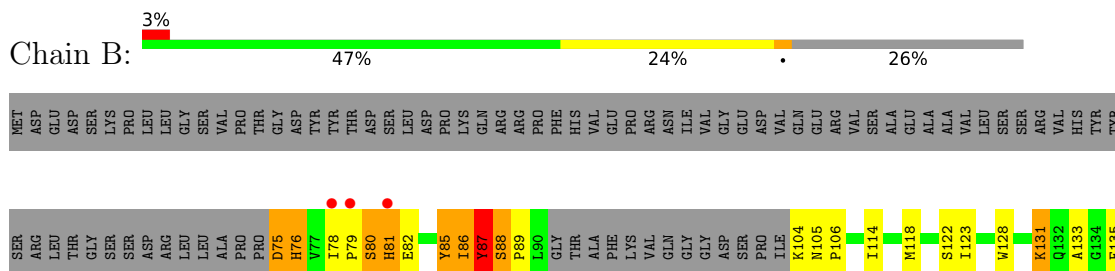
3 Residue-property plots

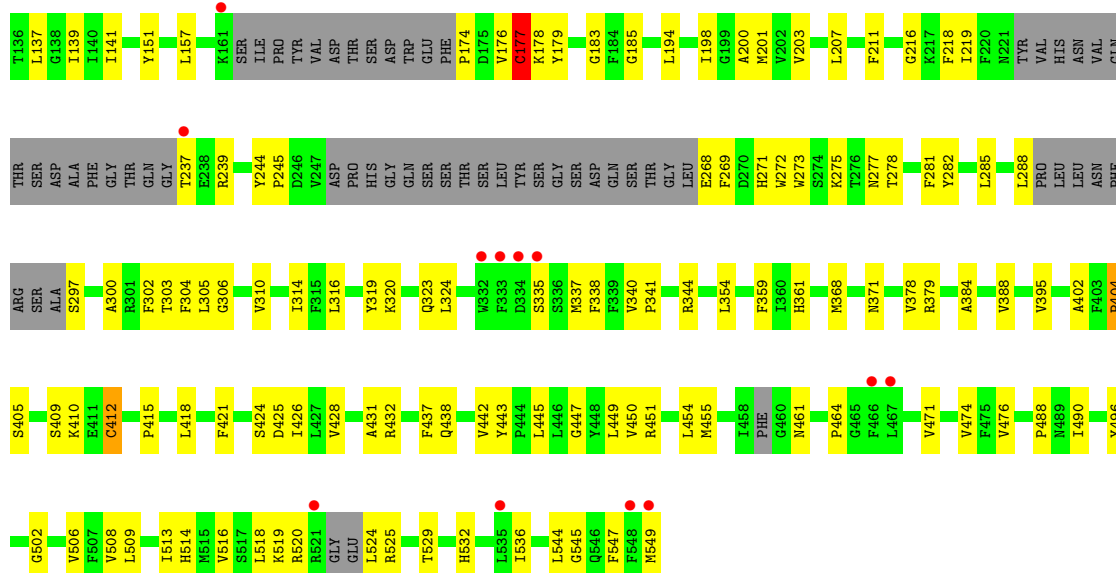
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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: Sodium-coupled neutral amino acid transporter 9

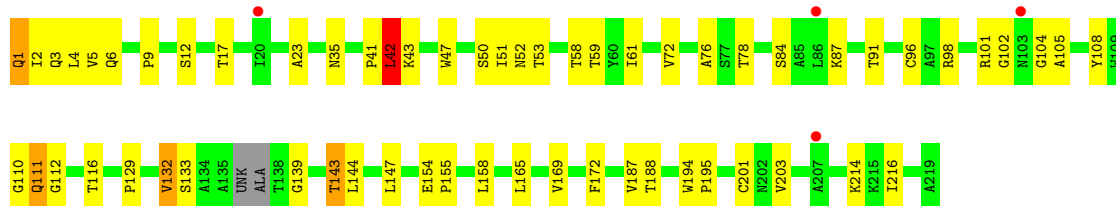


• Molecule 1: Sodium-coupled neutral amino acid transporter 9

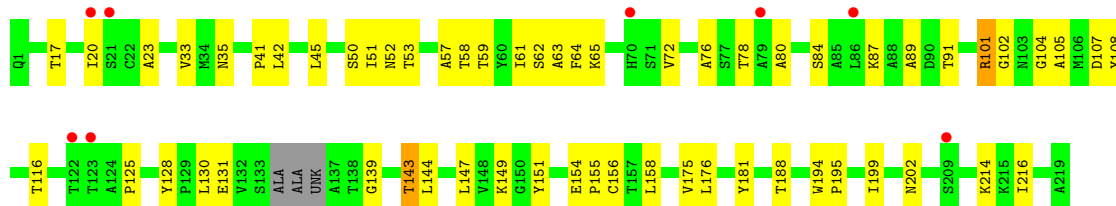
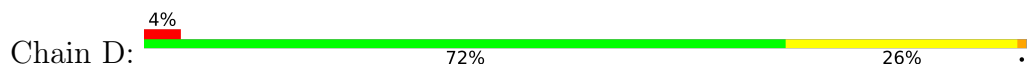




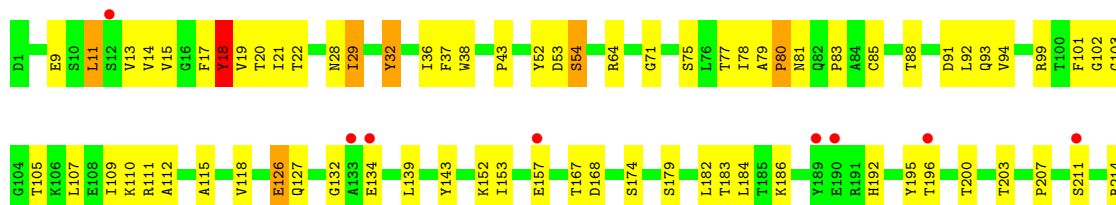
• Molecule 2: Monoclonal antibody Fab heavy chain



• Molecule 2: Monoclonal antibody Fab heavy chain



• Molecule 3: Monoclonal antibody Fab light chain



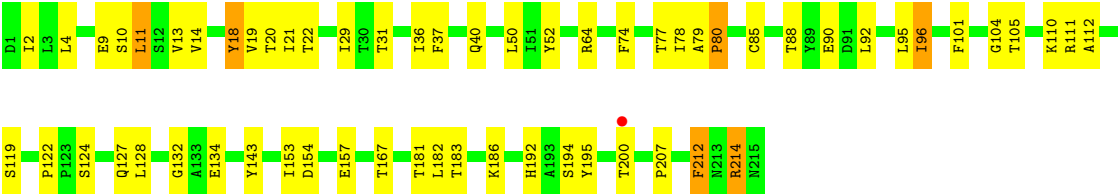
N215

• Molecule 3: Monoclonal antibody Fab light chain

Chain F:

72%

25%



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	157.80Å 82.51Å 158.59Å 90.00° 106.02° 90.00°	Depositor
Resolution (Å)	49.37 – 3.40 49.37 – 3.40	Depositor EDS
% Data completeness (in resolution range)	99.9 (49.37-3.40) 99.9 (49.37-3.40)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.36 (at 3.40Å)	Xtriage
Refinement program	PHENIX 1.18.2_3874+SVN	Depositor
R, R_{free}	0.251 , 0.284 0.251 , 0.284	Depositor DCC
R_{free} test set	1697 reflections (3.12%)	wwPDB-VP
Wilson B-factor (Å ²)	97.2	Xtriage
Anisotropy	0.357	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.23 , 48.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	0.024 for l,-k,h	Xtriage
F_o, F_c correlation	0.85	EDS
Total number of atoms	12613	wwPDB-VP
Average B, all atoms (Å ²)	110.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.61% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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 [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	0.43	0/2989	0.65	5/4059 (0.1%)
1	B	0.50	2/3367 (0.1%)	0.70	1/4575 (0.0%)
2	C	0.58	0/1610	0.76	0/2198
2	D	0.56	0/1605	0.75	1/2191 (0.0%)
3	E	0.60	0/1685	0.78	2/2296 (0.1%)
3	F	0.61	0/1685	0.74	1/2296 (0.0%)
All	All	0.53	2/12941 (0.0%)	0.72	10/17615 (0.1%)

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	B	0	1
2	C	0	1
All	All	0	2

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	177	CYS	CB-SG	-5.17	1.73	1.81
1	B	412	CYS	CB-SG	-5.09	1.73	1.81

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	286	LEU	CA-CB-CG	10.98	140.55	115.30
3	E	11	LEU	CB-CG-CD2	-8.63	96.33	111.00
1	A	286	LEU	CD1-CG-CD2	8.40	135.69	110.50
1	A	283	LEU	CA-CB-CG	7.27	132.03	115.30
3	F	11	LEU	CB-CG-CD2	-6.23	100.41	111.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	104	LYS	Peptide
2	C	1	GLN	Peptide

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	2905	0	2970	84	0
1	B	3269	0	3333	97	1
2	C	1574	0	1533	46	0
2	D	1569	0	1528	44	0
3	E	1648	0	1590	61	0
3	F	1648	0	1590	52	0
All	All	12613	0	12544	363	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 14.

The worst 5 of 363 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)
3:F:18:TYR:HE1	3:F:77:THR:HB	1.30	0.94
3:F:64:ARG:HB2	3:F:80:PRO:HD2	1.47	0.93
3:E:18:TYR:HE1	3:E:77:THR:HB	1.34	0.92
3:F:18:TYR:CE1	3:F:77:THR:HB	2.09	0.87
1:A:81:HIS:CD2	1:A:82:GLU:HG2	2.08	0.87

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)
1:B:461:ASN:O	1:B:520:ARG:NH2[2_656]	2.18	0.02

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	343/549 (62%)	302 (88%)	37 (11%)	4 (1%)	13	41
1	B	388/549 (71%)	336 (87%)	43 (11%)	9 (2%)	6	28
2	C	213/219 (97%)	196 (92%)	13 (6%)	4 (2%)	8	31
2	D	212/219 (97%)	195 (92%)	13 (6%)	4 (2%)	8	31
3	E	213/215 (99%)	199 (93%)	12 (6%)	2 (1%)	17	49
3	F	213/215 (99%)	200 (94%)	11 (5%)	2 (1%)	17	49
All	All	1582/1966 (80%)	1428 (90%)	129 (8%)	25 (2%)	9	34

5 of 25 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	87	TYR
1	B	88	SER
1	B	89	PRO
1	B	177	CYS
1	B	488	PRO

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	322/487 (66%)	306 (95%)	16 (5%)	24	54
1	B	362/487 (74%)	351 (97%)	11 (3%)	41	68
2	C	174/174 (100%)	168 (97%)	6 (3%)	37	65

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	D	174/174 (100%)	173 (99%)	1 (1%)	86	94
3	E	186/186 (100%)	179 (96%)	7 (4%)	33	61
3	F	186/186 (100%)	179 (96%)	7 (4%)	33	61
All	All	1404/1694 (83%)	1356 (97%)	48 (3%)	37	65

5 of 48 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	131	LYS
1	B	525	ARG
3	F	119	SER
1	B	268	GLU
1	B	378	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	349	GLN
1	B	438	GLN
2	C	6	GLN
2	C	35	ASN
3	F	127	GLN

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 monosaccharides 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](#)

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, 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(Å ²)	Q<0.9
1	A	361/549 (65%)	0.19	26 (7%) 15 17	93, 149, 193, 261	0
1	B	404/549 (73%)	-0.00	15 (3%) 41 40	70, 104, 160, 234	0
2	C	217/219 (99%)	-0.02	4 (1%) 68 67	62, 89, 134, 162	0
2	D	216/219 (98%)	0.02	8 (3%) 41 40	61, 86, 124, 158	0
3	E	215/215 (100%)	-0.01	8 (3%) 41 40	61, 85, 138, 160	0
3	F	215/215 (100%)	-0.20	1 (0%) 91 90	62, 89, 134, 184	0
All	All	1628/1966 (82%)	0.01	62 (3%) 40 39	61, 103, 175, 261	0

The worst 5 of 62 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	549	MET	6.3
1	A	81	HIS	5.0
1	A	516	VAL	4.6
1	B	334	ASP	4.6
1	A	80	SER	4.5

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 monosaccharides 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.