



wwPDB X-ray Structure Validation Summary Report ⓘ

May 21, 2020 – 04:50 pm BST

PDB ID : 5N7X
Title : CRYSTAL STRUCTURE OF STREPTAVIDIN WITH PEPTIDE EWWH-PQFEQKAK
Authors : Lyamichev, V.; Goodrich, L.; Sullivan, E.; Bannen, R.; Benz, J.; Albert, T.; Patel, J.
Deposited on : 2017-02-21
Resolution : 1.12 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	:	2.11
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

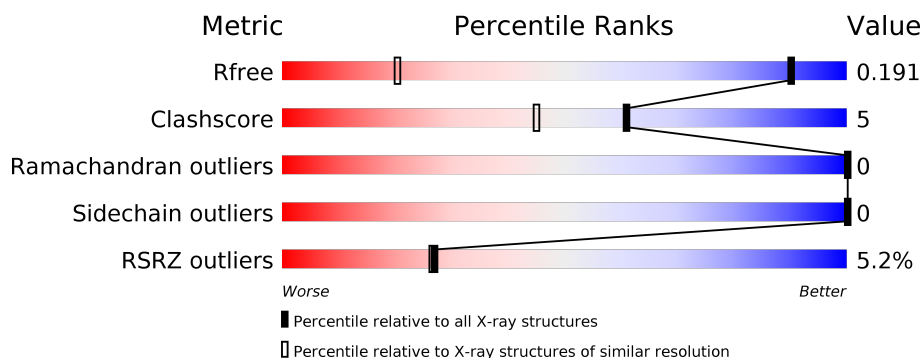
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.12 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	1168 (1.14-1.10)
Clashscore	141614	1205 (1.14-1.10)
Ramachandran outliers	138981	1168 (1.14-1.10)
Sidechain outliers	138945	1165 (1.14-1.10)
RSRZ outliers	127900	1146 (1.14-1.10)

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 $\leq 5\%$. 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	183	<div> <div>5%</div> <div> <div></div> <div>63%</div> <div></div> <div>33%</div> </div> </div>
1	B	183	<div> <div>3%</div> <div> <div></div> <div>62%</div> <div></div> <div>34%</div> </div> </div>
1	E	183	<div> <div>3%</div> <div> <div></div> <div>63%</div> <div></div> <div>33%</div> </div> </div>
1	F	183	<div> <div>3%</div> <div> <div></div> <div>63%</div> <div></div> <div>33%</div> </div> </div>
1	H	183	<div> <div>2%</div> <div> <div></div> <div>64%</div> <div></div> <div>33%</div> </div> </div>
1	K	183	<div> <div>3%</div> <div> <div></div> <div>63%</div> <div></div> <div>33%</div> </div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	M	183	<div><div></div><div>3%</div><div>61%</div><div>5%</div><div>34%</div></div>
1	O	183	<div><div></div><div>2%</div><div>64%</div><div></div><div>33%</div></div>
2	C	12	<div><div></div><div>17%</div><div>92%</div><div>8%</div></div>
2	D	12	<div><div></div><div>17%</div><div>92%</div><div>8%</div></div>
2	G	12	<div><div></div><div>8%</div><div>83%</div><div>8%</div><div>8%</div></div>
2	I	12	<div><div></div><div>8%</div><div>83%</div><div>8%</div><div>8%</div></div>
2	J	12	<div><div></div><div>8%</div><div>92%</div><div>8%</div></div>
2	L	12	<div><div></div><div>8%</div><div>92%</div><div>8%</div></div>
2	N	12	<div><div></div><div>8%</div><div>92%</div><div>8%</div></div>
2	P	12	<div><div></div><div>8%</div><div>92%</div><div>8%</div></div>

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 9634 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 Streptavidin.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
1	A	122	Total	C	N	O	0	10	0
			957	602	162	193			
1	B	121	Total	C	N	O	0	9	0
			945	596	160	189			
1	E	122	Total	C	N	O	0	8	0
			945	594	159	192			
1	F	122	Total	C	N	O	0	7	0
			939	590	158	191			
1	H	122	Total	C	N	O	6	9	0
			952	599	162	191			
1	K	122	Total	C	N	O	0	10	0
			951	599	158	194			
1	M	121	Total	C	N	O	0	8	0
			944	595	161	188			
1	O	122	Total	C	N	O	0	7	0
			939	590	158	191			

- Molecule 2 is a protein called GLU-TRP-VAL-HIS-PRO-GLN-PHE-GLU-GLN-LYS-ALA-LYS Peptide.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	C	11	Total	C	N	O	0	0	0
			99	65	17	17			
2	D	11	Total	C	N	O	0	0	0
			99	65	17	17			
2	G	11	Total	C	N	O	0	0	0
			99	65	17	17			
2	I	11	Total	C	N	O	0	0	0
			99	65	17	17			
2	J	11	Total	C	N	O	0	0	0
			99	65	17	17			
2	L	11	Total	C	N	O	0	0	0
			99	65	17	17			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	N	11	Total	C	N	O	0	0	0
			99	65	17	17			
2	P	11	Total	C	N	O	0	0	0
			99	65	17	17			

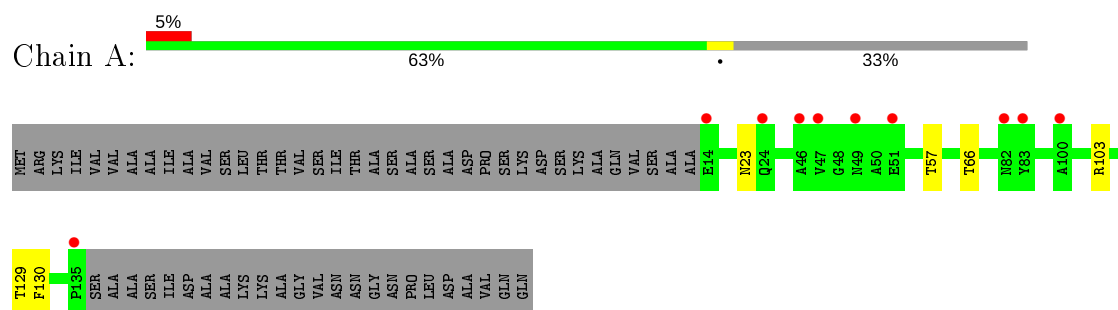
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	129	Total	O	0	0
			129	129		
3	B	153	Total	O	0	0
			153	153		
3	C	11	Total	O	0	0
			11	11		
3	D	9	Total	O	0	0
			9	9		
3	E	150	Total	O	0	0
			150	150		
3	F	140	Total	O	0	0
			140	140		
3	G	10	Total	O	0	0
			10	10		
3	H	148	Total	O	0	0
			148	148		
3	I	10	Total	O	0	0
			10	10		
3	J	24	Total	O	0	0
			24	24		
3	K	147	Total	O	0	0
			147	147		
3	L	14	Total	O	0	0
			14	14		
3	M	158	Total	O	0	0
			158	158		
3	N	11	Total	O	0	0
			11	11		
3	O	136	Total	O	0	0
			136	136		
3	P	20	Total	O	0	0
			20	20		

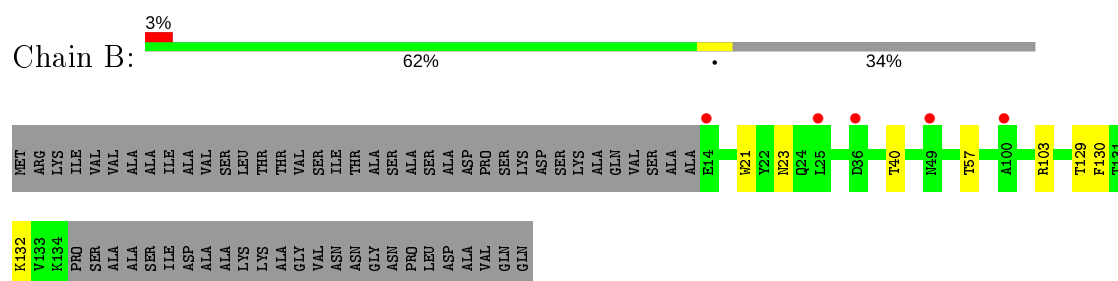
3 Residue-property plots [i](#)

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.

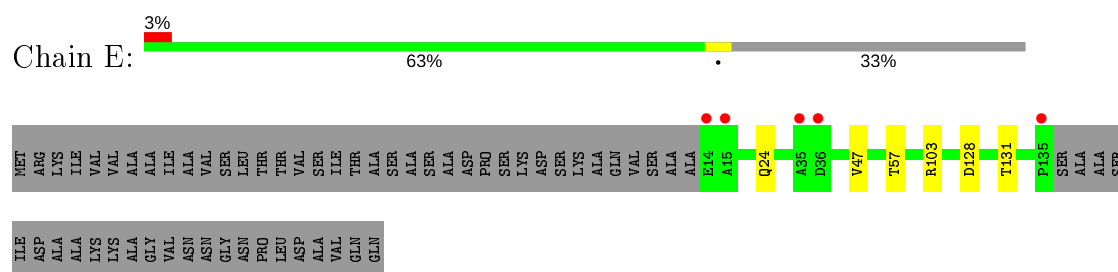
• Molecule 1: Streptavidin



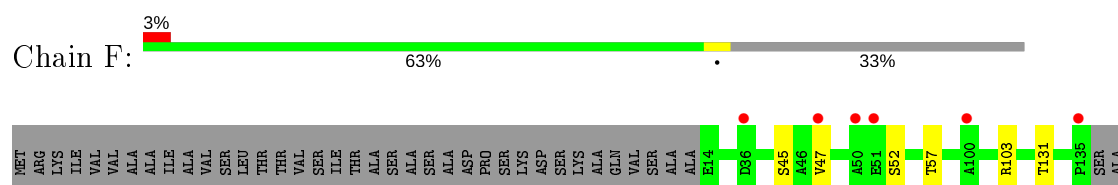
• Molecule 1: Streptavidin



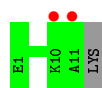
• Molecule 1: Streptavidin



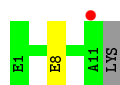
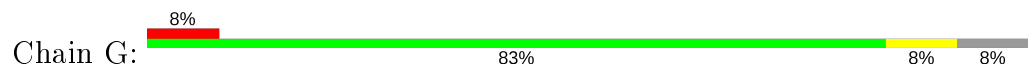
• Molecule 1: Streptavidin



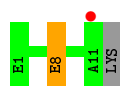
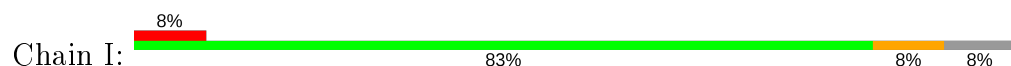
- Molecule 2: GLU-TRP-VAL-HIS-PRO-GLN-PHE-GLU-GLN-LYS-ALA-LYS Peptide



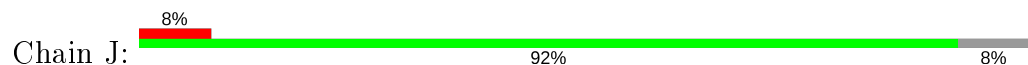
- Molecule 2: GLU-TRP-VAL-HIS-PRO-GLN-PHE-GLU-GLN-LYS-ALA-LYS Peptide



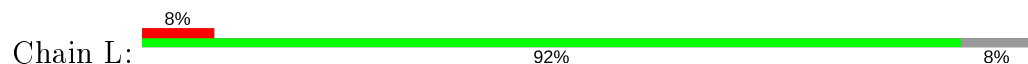
- Molecule 2: GLU-TRP-VAL-HIS-PRO-GLN-PHE-GLU-GLN-LYS-ALA-LYS Peptide



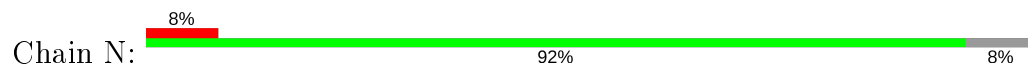
- Molecule 2: GLU-TRP-VAL-HIS-PRO-GLN-PHE-GLU-GLN-LYS-ALA-LYS Peptide



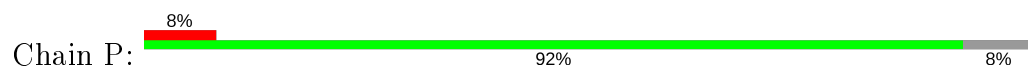
- Molecule 2: GLU-TRP-VAL-HIS-PRO-GLN-PHE-GLU-GLN-LYS-ALA-LYS Peptide



- Molecule 2: GLU-TRP-VAL-HIS-PRO-GLN-PHE-GLU-GLN-LYS-ALA-LYS Peptide



- Molecule 2: GLU-TRP-VAL-HIS-PRO-GLN-PHE-GLU-GLN-LYS-ALA-LYS Peptide





4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	57.91Å 64.81Å 81.16Å 96.31° 97.15° 84.09°	Depositor
Resolution (Å)	80.13 – 1.12 47.84 – 1.12	Depositor EDS
% Data completeness (in resolution range)	90.4 (80.13-1.12) 90.4 (47.84-1.12)	Depositor EDS
R_{merge}	0.03	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.59 (at 1.12Å)	Xtriage
Refinement program	REFMAC 5.8.0158	Depositor
R, R_{free}	0.160 , 0.186 0.169 , 0.191	Depositor DCC
R_{free} test set	20178 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å ²)	12.5	Xtriage
Anisotropy	0.134	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.36 , 47.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	9634	wwPDB-VP
Average B, all atoms (Å ²)	20.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 47.27 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.0159e-04.*

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	0.70	0/1008	0.88	0/1379
1	B	0.67	0/992	0.82	0/1357
1	E	0.71	0/990	0.81	0/1355
1	F	0.72	0/981	0.84	0/1343
1	H	0.76	1/1000 (0.1%)	0.84	0/1368
1	K	0.75	0/1002	0.86	0/1373
1	M	0.69	0/988	0.82	1/1351 (0.1%)
1	O	0.72	0/981	0.83	2/1343 (0.1%)
2	C	0.62	0/103	0.60	0/139
2	D	0.59	0/103	0.61	0/139
2	G	0.59	0/103	0.67	0/139
2	I	0.85	1/103 (1.0%)	0.80	0/139
2	J	0.62	0/103	0.61	0/139
2	L	0.64	0/103	0.63	0/139
2	N	0.57	0/103	0.59	0/139
2	P	0.66	0/103	0.64	0/139
All	All	0.71	2/8766 (0.0%)	0.82	3/11981 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	I	8	GLU	CD-OE1	5.40	1.31	1.25
1	H	135	PRO	C-O	-5.34	1.12	1.23

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	O	110	LEU	CB-CG-CD1	6.00	121.20	111.00
1	M	84	ARG	NE-CZ-NH1	-5.78	117.41	120.30
1	O	54	TYR	CB-CG-CD1	5.26	124.15	121.00

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	957	0	916	19	0
1	B	945	0	910	17	0
1	E	945	0	896	13	0
1	F	939	0	888	12	0
1	H	952	0	912	2	0
1	K	951	0	909	13	0
1	M	944	0	906	14	0
1	O	939	0	888	1	0
2	C	99	0	90	0	0
2	D	99	0	90	0	0
2	G	99	0	90	1	0
2	I	99	0	90	1	0
2	J	99	0	90	0	0
2	L	99	0	90	0	0
2	N	99	0	90	0	0
2	P	99	0	90	0	0
3	A	129	0	0	3	0
3	B	153	0	0	3	0
3	C	11	0	0	0	0
3	D	9	0	0	0	0
3	E	150	0	0	1	0
3	F	140	0	0	1	0
3	G	10	0	0	0	0
3	H	148	0	0	0	0
3	I	10	0	0	0	0
3	J	24	0	0	0	0
3	K	147	0	0	5	0
3	L	14	0	0	0	0
3	M	158	0	0	3	0
3	N	11	0	0	0	0
3	O	136	0	0	0	0
3	P	20	0	0	0	0
All	All	9634	0	7945	74	0

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 5.

The worst 5 of 74 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:M:42[B]:THR:HG21	1:M:53[B]:ARG:HD2	1.42	1.01
1:M:42[B]:THR:CG2	1:M:53[B]:ARG:CG	2.39	1.00
1:M:42[B]:THR:HG21	1:M:53[B]:ARG:CD	1.93	0.96
1:A:103:ARG:HH21	1:A:129[A]:THR:CG2	1.82	0.91
1:K:103:ARG:HH21	1:K:129[A]:THR:CG2	1.82	0.91

There are no symmetry-related clashes.

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	129/183 (70%)	128 (99%)	1 (1%)	0	100	100
1	B	127/183 (69%)	125 (98%)	2 (2%)	0	100	100
1	E	127/183 (69%)	125 (98%)	2 (2%)	0	100	100
1	F	126/183 (69%)	124 (98%)	2 (2%)	0	100	100
1	H	128/183 (70%)	125 (98%)	3 (2%)	0	100	100
1	K	129/183 (70%)	126 (98%)	3 (2%)	0	100	100
1	M	126/183 (69%)	125 (99%)	1 (1%)	0	100	100
1	O	126/183 (69%)	123 (98%)	3 (2%)	0	100	100
2	C	9/12 (75%)	9 (100%)	0	0	100	100
2	D	9/12 (75%)	9 (100%)	0	0	100	100
2	G	9/12 (75%)	9 (100%)	0	0	100	100
2	I	9/12 (75%)	9 (100%)	0	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	J	9/12 (75%)	9 (100%)	0	0	100	100
2	L	9/12 (75%)	9 (100%)	0	0	100	100
2	N	9/12 (75%)	9 (100%)	0	0	100	100
2	P	9/12 (75%)	9 (100%)	0	0	100	100
All	All	1090/1560 (70%)	1073 (98%)	17 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	99/134 (74%)	99 (100%)	0	100	100
1	B	97/134 (72%)	97 (100%)	0	100	100
1	E	97/134 (72%)	97 (100%)	0	100	100
1	F	96/134 (72%)	96 (100%)	0	100	100
1	H	98/134 (73%)	98 (100%)	0	100	100
1	K	99/134 (74%)	99 (100%)	0	100	100
1	M	96/134 (72%)	96 (100%)	0	100	100
1	O	96/134 (72%)	96 (100%)	0	100	100
2	C	10/11 (91%)	10 (100%)	0	100	100
2	D	10/11 (91%)	10 (100%)	0	100	100
2	G	10/11 (91%)	10 (100%)	0	100	100
2	I	10/11 (91%)	10 (100%)	0	100	100
2	J	10/11 (91%)	10 (100%)	0	100	100
2	L	10/11 (91%)	10 (100%)	0	100	100
2	N	10/11 (91%)	10 (100%)	0	100	100
2	P	10/11 (91%)	10 (100%)	0	100	100
All	All	858/1160 (74%)	858 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	F	82	ASN
1	O	82	ASN
1	K	82	ASN
1	E	82	ASN
1	H	82	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](#)

There are no chain breaks in this entry.

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, 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	122/183 (66%)	0.46	10 (8%) 11 12	11, 16, 35, 50	2 (1%)
1	B	121/183 (66%)	0.33	5 (4%) 37 35	10, 16, 33, 56	2 (1%)
1	E	122/183 (66%)	0.29	5 (4%) 37 35	10, 16, 30, 52	2 (1%)
1	F	122/183 (66%)	0.31	6 (4%) 29 28	9, 15, 29, 49	2 (1%)
1	H	122/183 (66%)	0.37	4 (3%) 46 45	9, 16, 30, 49	2 (1%)
1	K	122/183 (66%)	0.43	6 (4%) 29 28	10, 15, 33, 48	2 (1%)
1	M	121/183 (66%)	0.30	6 (4%) 28 28	9, 15, 31, 51	2 (1%)
1	O	122/183 (66%)	0.24	3 (2%) 57 55	9, 16, 29, 50	2 (1%)
2	C	11/12 (91%)	0.52	2 (18%) 1 2	13, 18, 50, 59	0
2	D	11/12 (91%)	1.23	2 (18%) 1 2	15, 26, 53, 58	0
2	G	11/12 (91%)	1.13	1 (9%) 9 10	12, 19, 39, 64	0
2	I	11/12 (91%)	1.29	1 (9%) 9 10	12, 18, 44, 65	0
2	J	11/12 (91%)	0.24	1 (9%) 9 10	12, 17, 35, 46	0
2	L	11/12 (91%)	0.48	1 (9%) 9 10	16, 23, 45, 50	0
2	N	11/12 (91%)	0.75	1 (9%) 9 10	15, 22, 49, 63	0
2	P	11/12 (91%)	0.39	1 (9%) 9 10	12, 17, 36, 49	0
All	All	1062/1560 (68%)	0.38	55 (5%) 27 26	9, 16, 34, 65	16 (1%)

The worst 5 of 55 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	I	11	ALA	12.4
2	G	11	ALA	10.7
1	K	135	PRO	7.7
2	D	11	ALA	6.1
1	A	49	ASN	5.8

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.