



wwPDB X-ray Structure Validation Summary Report ⓘ

May 14, 2020 – 11:39 am BST

PDB ID : 5MAV
Title : Crystal structure of human PCNA in complex with PARG (poly(ADP-ribose) glycohydrolase) peptide.
Authors : Grishkovskaya, I.; Djinovic-Carugo, K.; Slade, D.
Deposited on : 2016-11-06
Resolution : 2.58 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

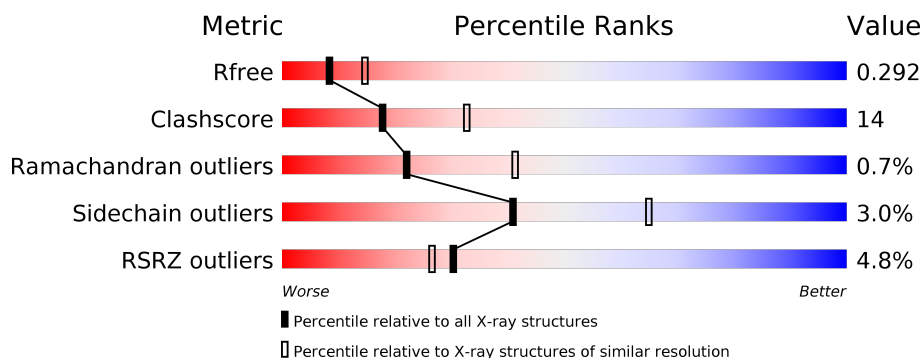
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.58 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	3676 (2.60-2.56)
Clashscore	141614	4049 (2.60-2.56)
Ramachandran outliers	138981	3979 (2.60-2.56)
Sidechain outliers	138945	3979 (2.60-2.56)
RSRZ outliers	127900	3614 (2.60-2.56)

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	264	<div> <div></div> <div> <div></div> <div>79%</div> <div>16%</div> <div>• •</div> </div> </div>
1	B	264	<div> <div>3%</div> <div> <div></div> <div>75%</div> <div>19%</div> <div>• •</div> </div> </div>
1	C	264	<div> <div></div> <div> <div></div> <div>80%</div> <div>13%</div> <div>• 5%</div> </div> </div>
1	D	264	<div> <div>14%</div> <div> <div></div> <div>51%</div> <div>38%</div> <div>• 8%</div> </div> </div>
1	E	264	<div> <div>2%</div> <div> <div></div> <div>76%</div> <div>17%</div> <div>• • 5%</div> </div> </div>
1	F	264	<div> <div>7%</div> <div> <div></div> <div>72%</div> <div>22%</div> <div>5%</div> </div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	G	19	
2	H	19	
2	K	19	
2	L	19	
2	M	19	
2	N	19	

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 12192 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 Proliferating cell nuclear antigen.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	254	Total	C	N	O	S	0	0	0
			1947	1224	319	388	16			
1	C	251	Total	C	N	O	S	0	2	0
			1942	1221	317	388	16			
1	E	251	Total	C	N	O	S	0	0	0
			1931	1214	316	385	16			
1	B	253	Total	C	N	O	S	0	0	0
			1941	1221	318	386	16			
1	D	243	Total	C	N	O	S	0	0	0
			1869	1177	307	370	15			
1	F	250	Total	C	N	O	S	0	0	0
			1922	1209	315	382	16			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	expression tag	UNP P12004
A	-1	PRO	-	expression tag	UNP P12004
A	0	ALA	-	expression tag	UNP P12004
C	-2	GLY	-	expression tag	UNP P12004
C	-1	PRO	-	expression tag	UNP P12004
C	0	ALA	-	expression tag	UNP P12004
E	-2	GLY	-	expression tag	UNP P12004
E	-1	PRO	-	expression tag	UNP P12004
E	0	ALA	-	expression tag	UNP P12004
B	-2	GLY	-	expression tag	UNP P12004
B	-1	PRO	-	expression tag	UNP P12004
B	0	ALA	-	expression tag	UNP P12004
D	-2	GLY	-	expression tag	UNP P12004
D	-1	PRO	-	expression tag	UNP P12004
D	0	ALA	-	expression tag	UNP P12004
F	-2	GLY	-	expression tag	UNP P12004
F	-1	PRO	-	expression tag	UNP P12004

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	0	ALA	-	expression tag	UNP P12004

- Molecule 2 is a protein called Poly (ADP-ribose) glycohydrolase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	G	14	Total	C	N	O	S	5	0	0
			119	76	22	20	1			
2	H	13	Total	C	N	O	S	0	0	0
			110	70	20	19	1			
2	K	12	Total	C	N	O	S	4	0	0
			103	65	19	18	1			
2	L	11	Total	C	N	O	S	0	0	0
			89	56	15	17	1			
2	N	10	Total	C	N	O	S	0	0	0
			84	53	14	16	1			
2	M	12	Total	C	N	O	S	0	0	0
			103	65	19	18	1			

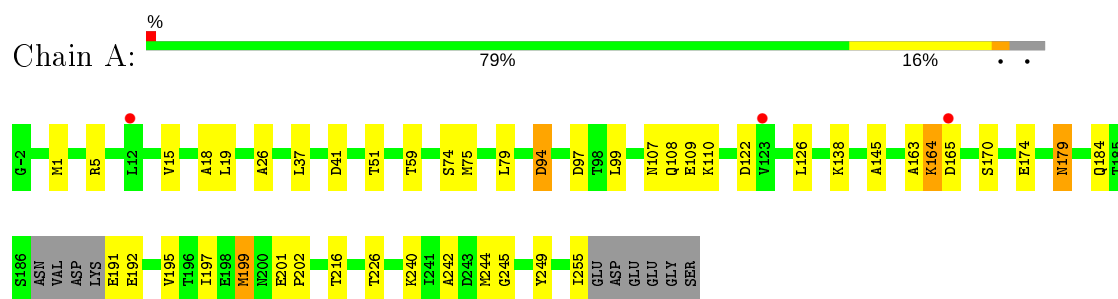
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	13	Total	O	0	0
			13	13		
3	C	7	Total	O	0	0
			7	7		
3	E	7	Total	O	0	0
			7	7		
3	B	2	Total	O	0	0
			2	2		
3	F	2	Total	O	0	0
			2	2		
3	H	1	Total	O	0	0
			1	1		

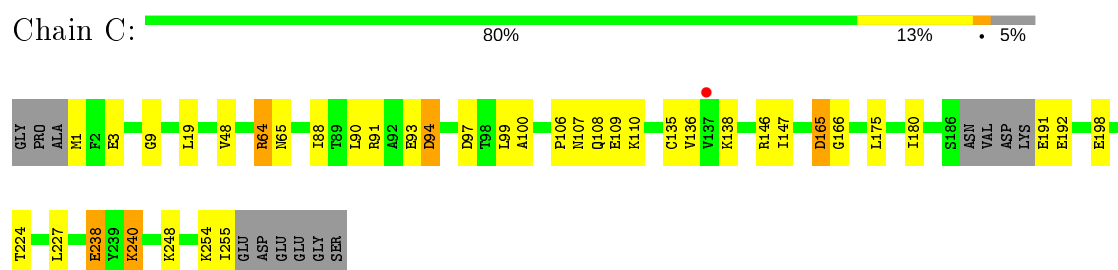
3 Residue-property plots [i](#)

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.

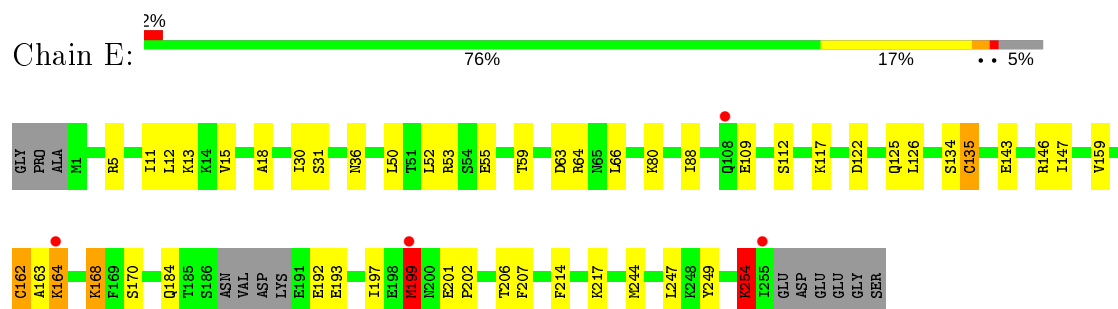
- Molecule 1: Proliferating cell nuclear antigen



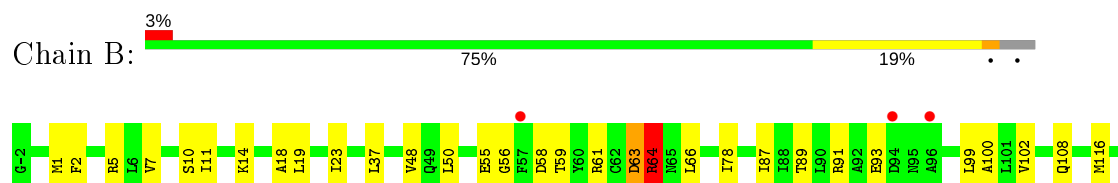
- Molecule 1: Proliferating cell nuclear antigen

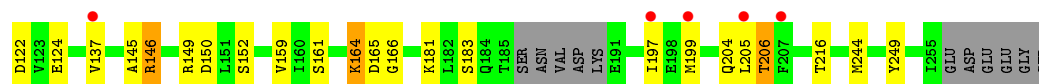


- Molecule 1: Proliferating cell nuclear antigen

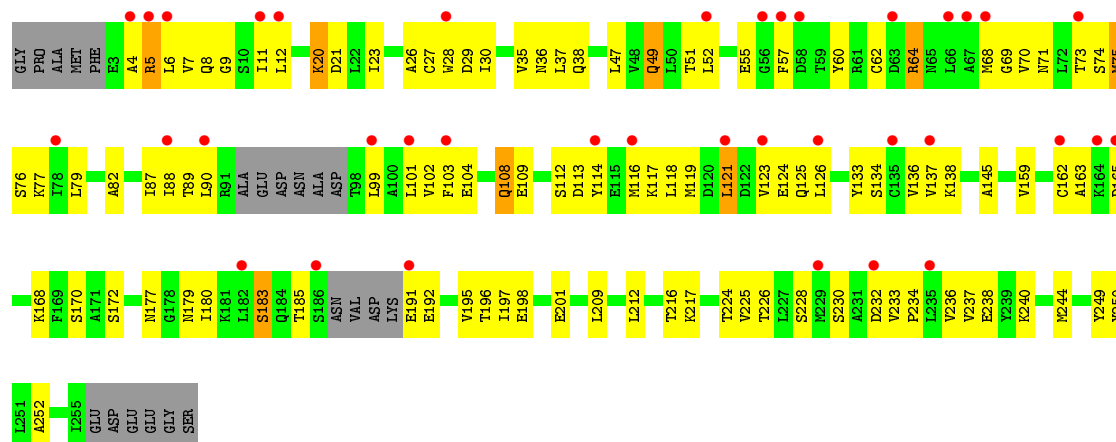


- Molecule 1: Proliferating cell nuclear antigen

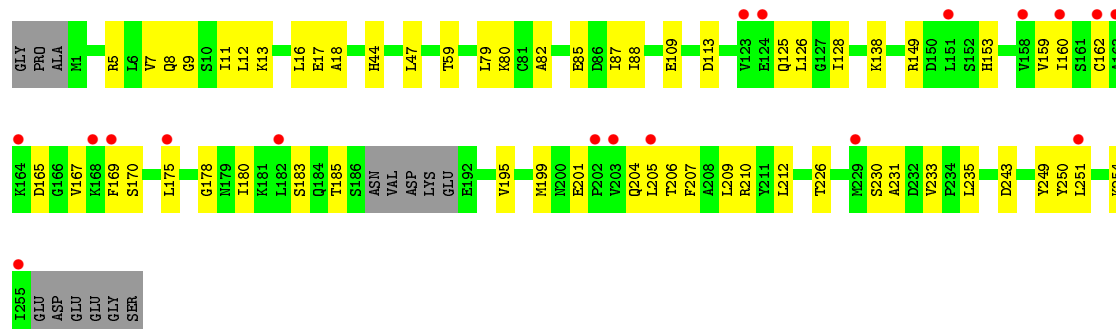




- Molecule 1: Proliferating cell nuclear antigen



- Molecule 1: Proliferating cell nuclear antigen



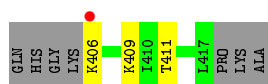
- Molecule 2: Poly (ADP-ribose) glycohydrolase



- Molecule 2: Poly (ADP-ribose) glycohydrolase



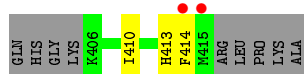
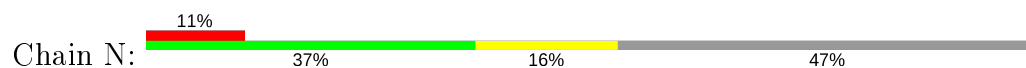
● Molecule 2: Poly (ADP-ribose) glycohydrolase



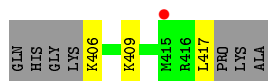
● Molecule 2: Poly (ADP-ribose) glycohydrolase



● Molecule 2: Poly (ADP-ribose) glycohydrolase



● Molecule 2: Poly (ADP-ribose) glycohydrolase



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	77.60 Å 153.60 Å 85.51 Å 90.00° 96.90° 90.00°	Depositor
Resolution (Å)	47.67 – 2.58 47.67 – 2.57	Depositor EDS
% Data completeness (in resolution range)	98.8 (47.67-2.58) 81.8 (47.67-2.57)	Depositor EDS
R_{merge}	0.26	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	0.38 (at 2.58 Å)	Xtriage
Refinement program	PHENIX (1.10 _2155: ???)	Depositor
R, R_{free}	0.239 , 0.291 0.240 , 0.292	Depositor DCC
R_{free} test set	3059 reflections (4.90%)	wwPDB-VP
Wilson B-factor (Å ²)	40.9	Xtriage
Anisotropy	0.649	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 58.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	12192	wwPDB-VP
Average B, all atoms (Å ²)	76.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.35% 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

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	0.28	0/1973	0.53	0/2665
1	B	0.31	0/1967	0.56	0/2657
1	C	0.29	0/1973	0.57	3/2664 (0.1%)
1	D	0.49	0/1892	0.79	6/2553 (0.2%)
1	E	0.39	0/1956	0.71	4/2641 (0.2%)
1	F	0.31	0/1947	0.61	0/2629
2	G	0.25	0/121	0.49	0/159
2	H	0.32	0/112	0.54	0/148
2	K	0.23	0/104	0.50	0/136
2	L	0.37	0/90	0.91	1/118 (0.8%)
2	M	0.41	0/104	0.74	0/136
2	N	0.45	0/85	0.74	0/111
All	All	0.35	0/12324	0.64	14/16617 (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	3

There are no bond length outliers.

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	E	254	LYS	CD-CE-NZ	-13.02	81.75	111.70
1	E	199	MET	CA-CB-CG	8.31	127.43	113.30
1	E	199	MET	CG-SD-CE	8.20	113.32	100.20
1	E	164	LYS	CA-CB-CG	7.19	129.21	113.40
1	D	121	LEU	CA-CB-CG	6.55	130.36	115.30

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	63	ASP	Peptide
1	B	64	ARG	Peptide
1	B	93	GLU	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	1947	0	1955	28	0
1	B	1941	0	1948	32	2
1	C	1942	0	1952	28	0
1	D	1869	0	1888	142	0
1	E	1931	0	1940	50	1
1	F	1922	0	1934	51	0
2	G	119	0	125	1	0
2	H	110	0	112	3	0
2	K	103	0	105	1	0
2	L	89	0	83	6	0
2	M	103	0	105	4	1
2	N	84	0	81	13	0
3	A	13	0	0	0	0
3	B	2	0	0	0	0
3	C	7	0	0	0	0
3	E	7	0	0	1	0
3	F	2	0	0	0	0
3	H	1	0	0	0	0
All	All	12192	0	12228	332	2

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 332 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:D:5:ARG:NH2	1:D:6:LEU:O	1.75	1.17

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:107:ASN:HD21	1:C:109:GLU:CD	1.50	1.14
1:C:107:ASN:ND2	1:C:109:GLU:OE1	1.85	1.09
1:E:207:PHE:HA	1:E:254:LYS:HZ1	0.98	1.09
1:D:118:LEU:HD23	1:D:119:MET:H	1.16	1.08

All (2) 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:55:GLU:O	2:M:409:LYS:NZ[1_655]	2.04	0.16
1:E:122:ASP:OD2	1:B:108:GLN:NE2[1_455]	2.13	0.07

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	250/264 (95%)	239 (96%)	8 (3%)	3 (1%)	13	26
1	B	249/264 (94%)	238 (96%)	10 (4%)	1 (0%)	34	55
1	C	249/264 (94%)	236 (95%)	11 (4%)	2 (1%)	19	37
1	D	237/264 (90%)	226 (95%)	9 (4%)	2 (1%)	19	37
1	E	247/264 (94%)	238 (96%)	9 (4%)	0	100	100
1	F	246/264 (93%)	233 (95%)	11 (4%)	2 (1%)	19	37
2	G	12/19 (63%)	11 (92%)	1 (8%)	0	100	100
2	H	11/19 (58%)	11 (100%)	0	0	100	100
2	K	10/19 (53%)	8 (80%)	2 (20%)	0	100	100
2	L	9/19 (47%)	9 (100%)	0	0	100	100
2	M	10/19 (53%)	10 (100%)	0	0	100	100
2	N	8/19 (42%)	8 (100%)	0	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	1538/1698 (91%)	1467 (95%)	61 (4%)	10 (1%)	22 41

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	94	ASP
1	C	165	ASP
1	B	64	ARG
1	D	165	ASP
1	A	244	MET

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	220/229 (96%)	212 (96%)	8 (4%)	35 59
1	B	219/229 (96%)	213 (97%)	6 (3%)	44 68
1	C	221/229 (96%)	215 (97%)	6 (3%)	44 68
1	D	213/229 (93%)	206 (97%)	7 (3%)	38 61
1	E	219/229 (96%)	211 (96%)	8 (4%)	34 57
1	F	218/229 (95%)	214 (98%)	4 (2%)	59 78
2	G	14/17 (82%)	12 (86%)	2 (14%)	3 5
2	H	13/17 (76%)	13 (100%)	0	100 100
2	K	12/17 (71%)	11 (92%)	1 (8%)	11 21
2	L	10/17 (59%)	10 (100%)	0	100 100
2	M	12/17 (71%)	12 (100%)	0	100 100
2	N	10/17 (59%)	10 (100%)	0	100 100
All	All	1381/1476 (94%)	1339 (97%)	42 (3%)	41 65

5 of 42 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	162	CYS
1	B	146	ARG
1	F	210	ARG
1	E	168	LYS
1	E	254	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	108	GLN
1	E	184	GLN
1	D	246	HIS
1	E	49	GLN
1	F	24	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 [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	254/264 (96%)	0.18	3 (1%) 79 77	33, 51, 93, 122	0
1	B	253/264 (95%)	0.40	8 (3%) 47 43	46, 75, 115, 137	0
1	C	251/264 (95%)	0.19	1 (0%) 92 92	35, 61, 97, 130	0
1	D	243/264 (92%)	0.96	37 (15%) 2 1	64, 101, 137, 149	0
1	E	251/264 (95%)	0.24	4 (1%) 72 69	42, 66, 109, 142	0
1	F	250/264 (94%)	0.50	18 (7%) 15 13	48, 80, 115, 149	0
2	G	14/19 (73%)	0.04	0 100 100	51, 65, 79, 86	1 (7%)
2	H	13/19 (68%)	0.07	0 100 100	46, 55, 73, 82	0
2	K	12/19 (63%)	0.59	1 (8%) 11 9	77, 88, 100, 103	1 (8%)
2	L	11/19 (57%)	0.19	0 100 100	82, 95, 120, 123	0
2	M	12/19 (63%)	0.23	1 (8%) 11 9	65, 83, 100, 100	0
2	N	10/19 (52%)	0.97	2 (20%) 1 0	89, 95, 101, 103	0
All	All	1574/1698 (92%)	0.40	75 (4%) 30 26	33, 72, 121, 149	2 (0%)

The worst 5 of 75 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	164	LYS	6.3
1	D	56	GLY	5.2
1	D	12	LEU	4.9
1	A	123	VAL	4.6
1	D	67	ALA	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 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.