



Full wwPDB X-ray Structure Validation Report ⓘ

Aug 22, 2022 – 11:19 AM EDT

PDB ID : 8DF2
Title : The structure of the 'ALT' construct of the Amuc_1438 glycopeptidase
Authors : Medley, B.J.; Boraston, A.B.
Deposited on : 2022-06-21
Resolution : 2.35 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 types of validation reports are described at <http://www.wwpdb.org/validation/2017/FAQs#types>.

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 : 2.29
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) : 2.29

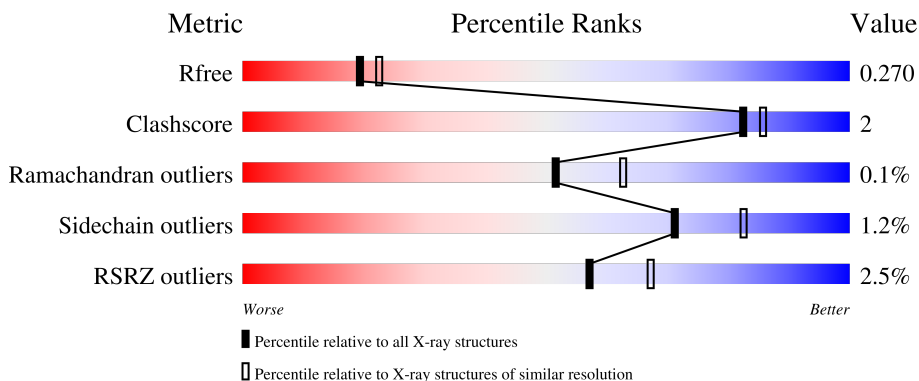
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.35 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	1164 (2.36-2.36)
Clashscore	141614	1232 (2.36-2.36)
Ramachandran outliers	138981	1211 (2.36-2.36)
Sidechain outliers	138945	1212 (2.36-2.36)
RSRZ outliers	127900	1150 (2.36-2.36)

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	486	
1	B	486	
1	C	486	
1	D	486	

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 11578 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 NPCBM/NEW2 domain-containing protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	374	2798	1773	483	529	13	0	0	0
1	B	374	2816	1784	487	532	13	0	0	0
1	C	374	2807	1778	487	529	13	0	1	0
1	D	374	2827	1790	491	532	14	0	0	0

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total 1	Zn 1	0	0
2	B	1	Total 1	Zn 1	0	0
2	C	1	Total 1	Zn 1	0	0
2	D	1	Total 1	Zn 1	0	0

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total 1	Ca 1	0	0
3	B	1	Total 1	Ca 1	0	0
3	C	1	Total 1	Ca 1	0	0
3	D	1	Total 1	Ca 1	0	0

- Molecule 4 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	3	Total 3	Na 3	0	0
4	B	4	Total 4	Na 4	0	0
4	C	4	Total 4	Na 4	0	0
4	D	1	Total 1	Na 1	0	0

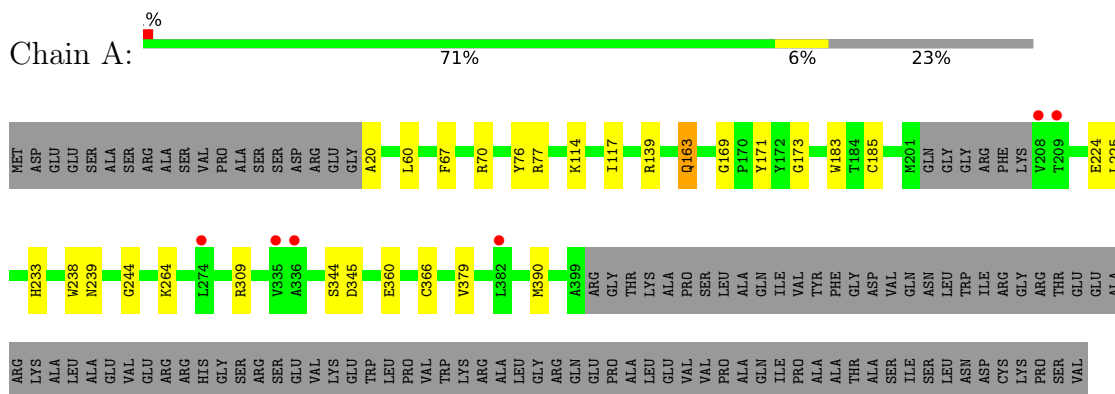
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	79	Total 79	O 79	0	0
5	B	89	Total 89	O 89	0	0
5	C	68	Total 68	O 68	0	0
5	D	74	Total 74	O 74	0	0

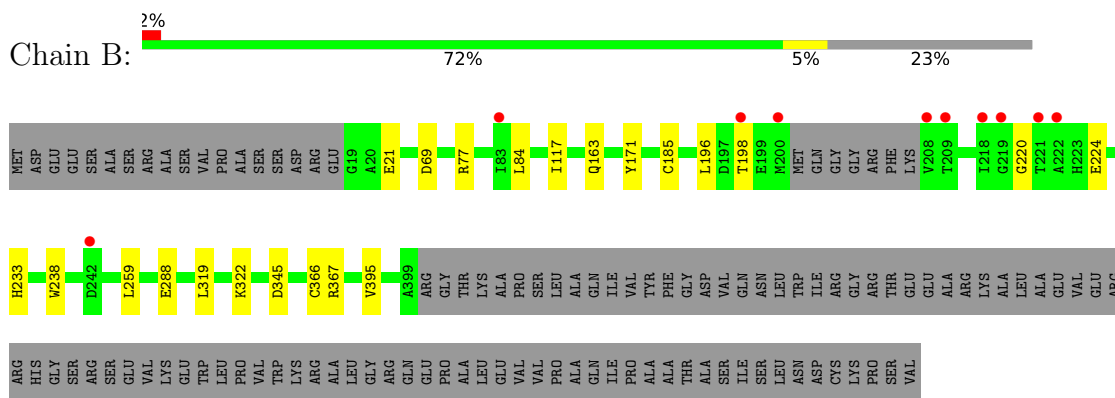
3 Residue-property plots [i](#)

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.

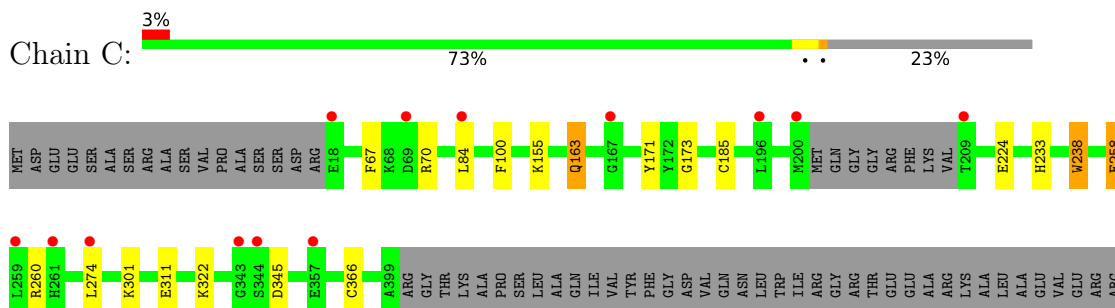
- Molecule 1: NPCBM/NEW2 domain-containing protein



- Molecule 1: NPCBM/NEW2 domain-containing protein



- Molecule 1: NPCBM/NEW2 domain-containing protein



HIS
GLY
SER
ARG
SER
GLU
VAL
VAL
LYS
GLU
TRP
LEU
PRO
VAL
TRP
LYS
ARG
ALA
LEU
GLY
ARG
GLN
GLU
PRO
ALA
LEU
GLU
VAL
VAL
PRO
ALA
GLN
ILE
PRO
ALA
ALA
THR
SER
ILE
LEU
ASN
ASP
CYS
LYS
PRO
SER
VAL

- Molecule 1: NPCBM/NEW2 domain-containing protein

Chain D: 2% 71% 5% 23%

MET
ASP
GLU
SER
SER
ALA
SER
ARG
ALA
SER
VAL
PRO
ALA
SER
SER
ASP
ARG
GLU
GLY
K20
K63
F67
R70
R77
E78
R79
Y90
D119
Q132
D152
K155
Q163
Y172
L190
M201
GLN
GLY
GLY
ARG
PHE
LYS
V208
K222
H233

W238
D242
L247
E258
L289
R290
D272
A273
L274
T287
R309
V335
A336
H337
D345
L355
E360
F361
M377
A399
ARG
GLY
THR
LYS
ALA
PRO
SER
LEU
ALA
GLN
ILE
VAL
TYR
PHE
GLY
ASP
VAL
GLN
ASN
LEU
TRP
ILE
ARG
GLY
ARG
THR
GLU

ALA
ARG
LYS
LEU
GLU
VAL
GLU
ARG
HIS
GLY
SER
ARG
SER
GLU
VAL
LYS
GLU
TRP
LEU
PRO
VAL
TRP
LYS
ARG
ALA
LEU
GLY
ARG
GLN
GLU
PRO
ALA
LEU
GLU
VAL
VAL
PRO
ALA
GLN
ILE
PRO
ALA
ALA
THR
THR
ALA
SER
ILE
SER
LEU
ASN
ASP
CYS
LYS
PRO
SER
VAL

4 Data and refinement statistics i

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	88.58Å 146.09Å 147.29Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	24.93 – 2.35 24.93 – 2.35	Depositor EDS
% Data completeness (in resolution range)	95.8 (24.93-2.35) 95.8 (24.93-2.35)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.90 (at 2.36Å)	Xtrriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, R_{free}	0.222 , 0.270 0.222 , 0.270	Depositor DCC
R_{free} test set	3848 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å ²)	35.7	Xtrriage
Anisotropy	0.290	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 24.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	0.000 for -h,l,k	Xtrriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	11578	wwPDB-VP
Average B, all atoms (Å ²)	33.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.98% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, CA, NA

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	0.32	0/2870	0.54	0/3899
1	B	0.33	0/2888	0.54	0/3920
1	C	0.32	0/2879	0.55	0/3902
1	D	0.31	0/2900	0.54	0/3939
All	All	0.32	0/11537	0.54	0/15660

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	2798	0	2637	18	0
1	B	2816	0	2677	12	0
1	C	2807	0	2628	13	0
1	D	2827	0	2695	12	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
3	A	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	1	0	0	0	0
3	C	1	0	0	0	0
3	D	1	0	0	0	0
4	A	3	0	0	0	0
4	B	4	0	0	0	0
4	C	4	0	0	0	0
4	D	1	0	0	0	0
5	A	79	0	0	2	0
5	B	89	0	0	1	0
5	C	68	0	0	1	0
5	D	74	0	0	0	0
All	All	11578	0	10637	54	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 2.

All (54) 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:366:CYS:SG	5:A:667:HOH:O	2.39	0.80
1:B:319:LEU:HD11	1:B:395:VAL:HG11	1.65	0.77
1:D:77:ARG:NH2	1:D:119:ASP:OD2	2.34	0.59
1:C:258:GLU:C	1:C:260:ARG:H	2.07	0.57
1:B:366:CYS:SG	5:B:684:HOH:O	2.48	0.55
1:C:100:PHE:HZ	1:C:274:LEU:HD23	1.72	0.55
1:B:288:GLU:OE1	1:C:155:LYS:NZ	2.37	0.55
1:D:258:GLU:O	1:D:260:ARG:N	2.41	0.54
1:D:309:ARG:NH2	1:D:360:GLU:OE1	2.42	0.53
1:A:20:ALA:N	5:A:602:HOH:O	2.42	0.53
1:D:67:PHE:HB2	1:D:70:ARG:HB2	1.92	0.51
1:A:76:TYR:OH	1:A:77:ARG:NH1	2.45	0.49
1:A:169:GLY:HA3	1:A:185:CYS:SG	2.53	0.49
1:C:70:ARG:HB3	1:C:163:GLN:NE2	2.28	0.49
1:B:69:ASP:OD1	1:B:69:ASP:N	2.44	0.48
1:A:344:SER:OG	1:A:345:ASP:N	2.47	0.48
1:D:152:ASP:HB3	1:D:155:LYS:HG2	1.94	0.48
1:C:173:GLY:N	1:C:224:GLU:HG2	2.29	0.47
1:B:77:ARG:HA	1:B:117:ILE:HD12	1.96	0.47
1:C:67:PHE:HB2	1:C:70:ARG:HB2	1.97	0.47
1:D:90:TYR:HE2	1:D:247:LEU:HD22	1.80	0.47
1:A:60:LEU:HD11	1:A:225:LEU:HD22	1.97	0.47

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:21:GLU:H	1:B:21:GLU:CD	2.19	0.46
1:A:77:ARG:HA	1:A:117:ILE:HD12	1.98	0.45
1:D:355:LEU:HD23	1:D:361:PHE:HB3	1.98	0.45
1:C:258:GLU:C	1:C:260:ARG:N	2.70	0.45
1:A:171:TYR:OH	1:A:185:CYS:HB2	2.17	0.45
1:D:337:HIS:O	1:D:377:MET:HA	2.18	0.44
1:A:239:ASN:OD1	1:A:239:ASN:N	2.51	0.44
1:B:220:GLY:O	1:B:224:GLU:HG2	2.17	0.44
1:A:233:HIS:HB2	1:A:345:ASP:CG	2.38	0.44
1:A:114:LYS:HD3	1:A:114:LYS:HA	1.89	0.43
1:A:173:GLY:N	1:A:224:GLU:HG2	2.33	0.43
1:D:233:HIS:HB2	1:D:345:ASP:CG	2.39	0.43
1:A:139:ARG:HB2	1:A:183:TRP:HZ2	1.84	0.43
1:A:67:PHE:HB2	1:A:70:ARG:HB2	2.01	0.43
1:A:70:ARG:HH21	1:A:163:GLN:HE22	1.67	0.42
1:B:196:LEU:HD23	1:B:259:LEU:HB3	2.02	0.42
1:B:171:TYR:OH	1:B:185:CYS:HB2	2.20	0.42
1:C:311:GLU:OE1	1:C:322:LYS:HE2	2.19	0.42
1:A:244:GLY:HA2	1:A:264:LYS:HB3	2.02	0.42
1:D:70:ARG:HH21	1:D:163:GLN:HE22	1.68	0.42
1:C:233:HIS:HB2	1:C:345:ASP:CG	2.40	0.41
1:A:379:VAL:O	1:A:390:MET:HA	2.20	0.41
1:C:84:LEU:HA	1:C:84:LEU:HD13	1.87	0.41
1:B:322:LYS:HB2	1:B:322:LYS:HE2	1.65	0.41
1:A:309:ARG:NH2	1:A:360:GLU:OE1	2.54	0.41
1:B:84:LEU:HD13	1:B:84:LEU:HA	1.88	0.41
1:B:233:HIS:HB2	1:B:345:ASP:CG	2.41	0.41
1:C:238:TRP:CE2	1:C:366:CYS:HB2	2.55	0.41
1:D:53:LYS:HD2	1:D:287:THR:HG22	2.02	0.41
1:D:79:ARG:HD2	1:D:190:LEU:O	2.20	0.41
1:C:171:TYR:CZ	1:C:185:CYS:HB2	2.56	0.40
1:C:301:LYS:NZ	5:C:603:HOH:O	2.42	0.40

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	370/486 (76%)	358 (97%)	12 (3%)	0	100	100
1	B	370/486 (76%)	361 (98%)	9 (2%)	0	100	100
1	C	371/486 (76%)	360 (97%)	11 (3%)	0	100	100
1	D	370/486 (76%)	359 (97%)	10 (3%)	1 (0%)	41	47
All	All	1481/1944 (76%)	1438 (97%)	42 (3%)	1 (0%)	51	63

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	259	LEU

5.3.2 Protein sidechains [i](#)

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	278/390 (71%)	276 (99%)	2 (1%)	84	91
1	B	280/390 (72%)	276 (99%)	4 (1%)	67	78
1	C	273/390 (70%)	270 (99%)	3 (1%)	73	84
1	D	285/390 (73%)	281 (99%)	4 (1%)	67	78
All	All	1116/1560 (72%)	1103 (99%)	13 (1%)	71	82

All (13) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	163	GLN
1	A	238	TRP
1	B	163	GLN
1	B	198	THR
1	B	238	TRP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	367	ARG
1	C	163	GLN
1	C	238	TRP
1	C	258	GLU
1	D	132	GLN
1	D	163	GLN
1	D	208	VAL
1	D	238	TRP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	163	GLN
1	D	132	GLN
1	D	163	GLN

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 monosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 20 ligands modelled in this entry, 20 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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	374/486 (76%)	0.04	6 (1%) 72 80	24, 31, 42, 57	0
1	B	374/486 (76%)	0.07	10 (2%) 54 64	24, 31, 46, 64	0
1	C	374/486 (76%)	0.12	13 (3%) 44 56	25, 35, 50, 65	0
1	D	374/486 (76%)	0.07	8 (2%) 63 74	25, 33, 42, 49	0
All	All	1496/1944 (76%)	0.08	37 (2%) 57 67	24, 32, 46, 65	0

All (37) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	208	VAL	4.0
1	C	209	THR	3.7
1	A	208	VAL	3.3
1	C	261	HIS	3.2
1	B	198	THR	3.1
1	A	209	THR	2.9
1	D	336	ALA	2.9
1	D	274	LEU	2.7
1	C	167	GLY	2.7
1	A	335	VAL	2.6
1	D	172	TYR	2.5
1	D	273	ALA	2.5
1	C	259	LEU	2.5
1	D	335	VAL	2.5
1	D	222	ALA	2.4
1	C	69	ASP	2.3
1	B	200	MET	2.3
1	B	242	ASP	2.3
1	B	83	ILE	2.3
1	B	209	THR	2.3
1	C	344	SER	2.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	C	84	LEU	2.3
1	B	218	ILE	2.3
1	C	200	MET	2.2
1	C	18	GLU	2.2
1	B	222	ALA	2.1
1	D	242	ASP	2.1
1	C	274	LEU	2.1
1	A	382	LEU	2.1
1	C	357	GLU	2.1
1	D	272	ASP	2.1
1	A	336	ALA	2.1
1	A	274	LEU	2.1
1	B	221	THR	2.1
1	B	219	GLY	2.0
1	C	343	GLY	2.0
1	C	196	LEU	2.0

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](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å ²)	Q<0.9
4	NA	A	505	1/1	0.75	0.20	50,50,50,50	0
4	NA	B	504	1/1	0.80	0.14	47,47,47,47	0
4	NA	B	506	1/1	0.81	0.07	48,48,48,48	0
4	NA	D	503	1/1	0.81	0.13	44,44,44,44	0
4	NA	B	505	1/1	0.89	0.15	53,53,53,53	0
4	NA	C	506	1/1	0.90	0.09	41,41,41,41	0
4	NA	A	503	1/1	0.93	0.07	50,50,50,50	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
4	NA	C	503	1/1	0.93	0.15	52,52,52,52	0
4	NA	A	504	1/1	0.94	0.14	45,45,45,45	0
4	NA	C	504	1/1	0.94	0.10	48,48,48,48	0
4	NA	B	503	1/1	0.95	0.10	40,40,40,40	0
3	CA	B	502	1/1	0.95	0.05	32,32,32,32	0
4	NA	C	505	1/1	0.97	0.15	20,20,20,20	0
3	CA	A	502	1/1	0.98	0.06	31,31,31,31	0
3	CA	C	502	1/1	0.98	0.04	34,34,34,34	0
3	CA	D	502	1/1	0.98	0.06	36,36,36,36	0
2	ZN	A	501	1/1	0.99	0.07	35,35,35,35	0
2	ZN	B	501	1/1	0.99	0.07	34,34,34,34	0
2	ZN	C	501	1/1	0.99	0.04	39,39,39,39	0
2	ZN	D	501	1/1	0.99	0.03	36,36,36,36	0

6.5 Other polymers [i](#)

There are no such residues in this entry.