



wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 8, 2023 – 10:13 AM JST

PDB ID : 7VYX
Title : Crystal structure of the selenomethionine(SeMet)-derived Cas12c1 (D969A) ternary complex
Authors : Zhang, B.; Lin, J.Y.; Perculija, V.; OuYang, S.Y.
Deposited on : 2021-11-15
Resolution : 3.20 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 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
Mogul : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.33
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) : 2.33

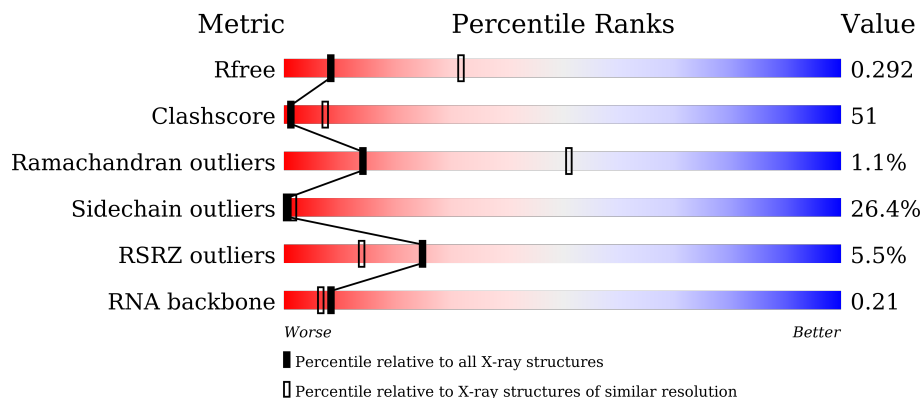
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.20 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.



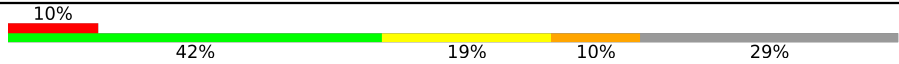

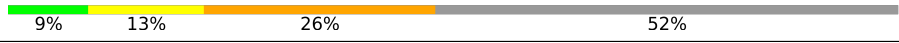

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1133 (3.20-3.20)
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)
RSRZ outliers	127900	1095 (3.20-3.20)
RNA backbone	3102	1010 (3.50-2.90)

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	1310	
1	B	1310	
2	C	136	
2	D	136	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	E	31	
3	F	31	
4	G	23	
4	H	23	

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 24465 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 Selenomethionine (SeMet)-labeled Cas12c1 D969A mutant.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	1211	Total 9533	C 6075	N 1639	O 1788	S 20	Se 11	0	0	0
1	B	1193	Total 9369	C 5977	N 1607	O 1754	S 20	Se 11	0	0	0

- Molecule 2 is a RNA chain called sgRNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	C	103	Total 2188	C 978	N 387	O 720	P 103	0	0	0
2	D	91	Total 1934	C 865	N 344	O 634	P 91	0	0	0

- Molecule 3 is a DNA chain called Target DNA strand.

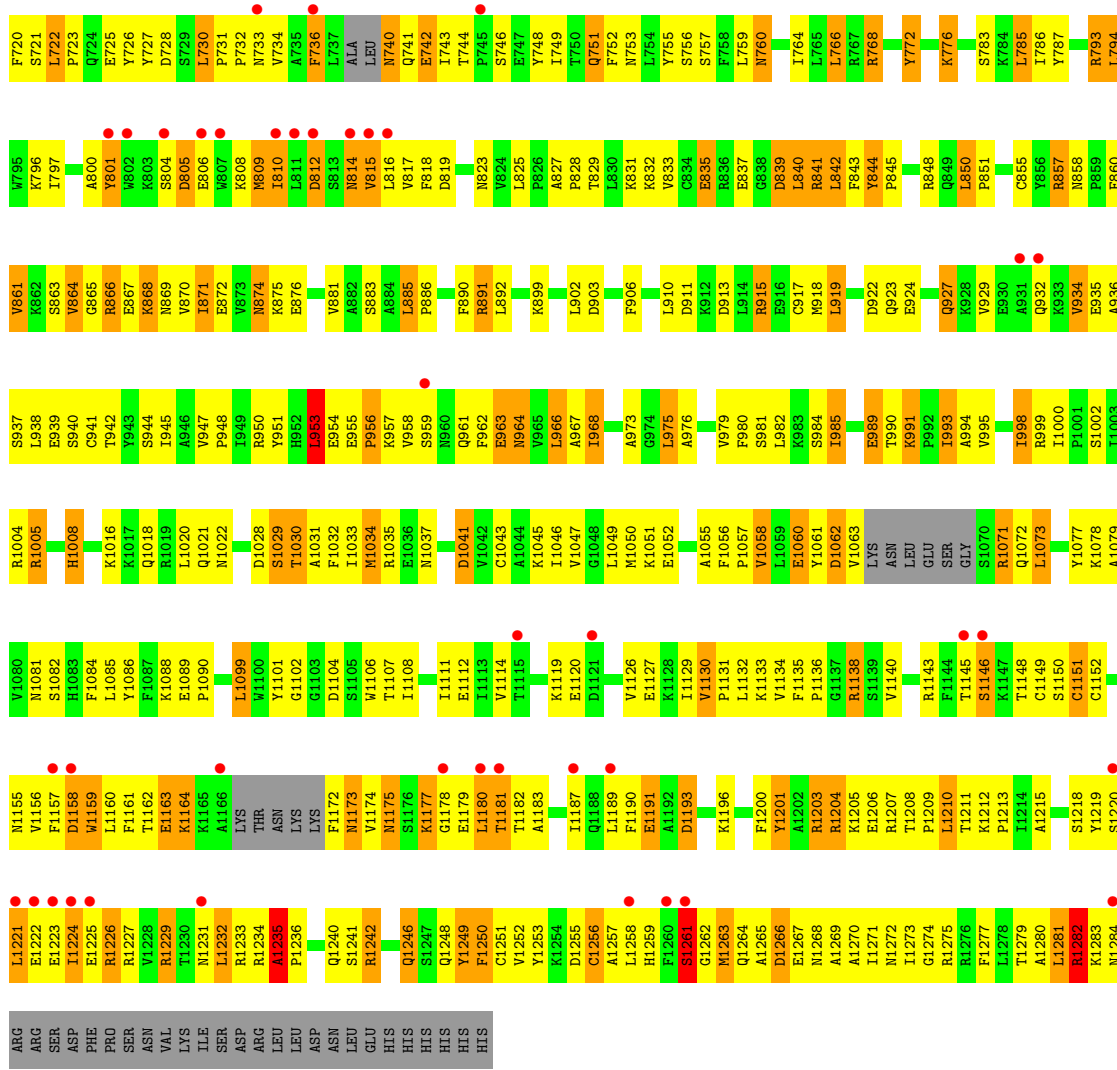
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	E	22	Total 451	C 216	N 78	O 135	P 22	0	0	0
3	F	26	Total 534	C 256	N 92	O 160	P 26	0	0	0

- Molecule 4 is a DNA chain called Non-target DNA strand.

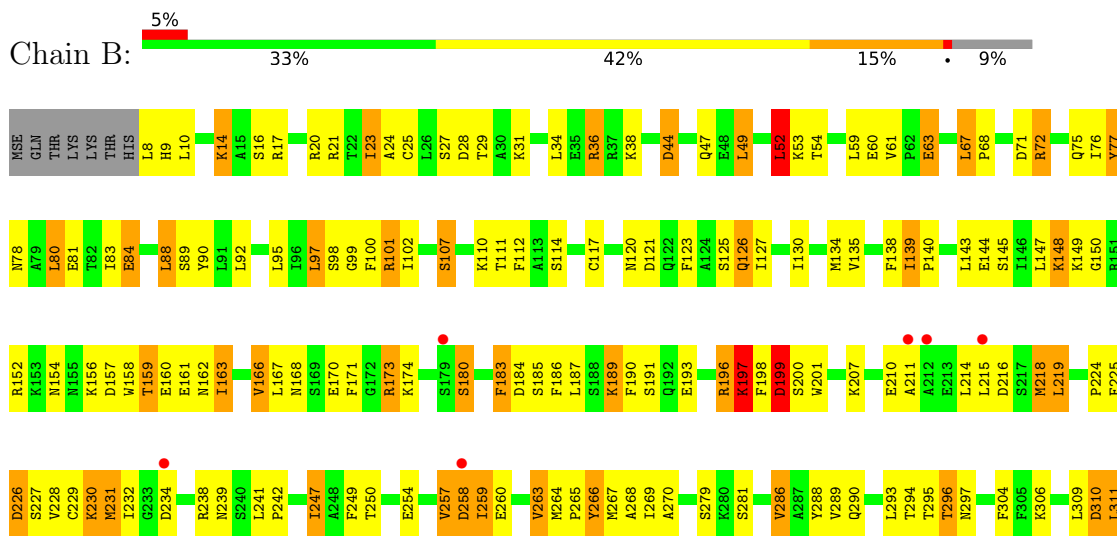
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
4	G	11	Total 227	C 108	N 42	O 66	P 11	0	0	0
4	H	11	Total 227	C 108	N 42	O 66	P 11	0	0	0

- Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	1	Total 1	Zn 1	0	0
5	B	1	Total 1	Zn 1	0	0

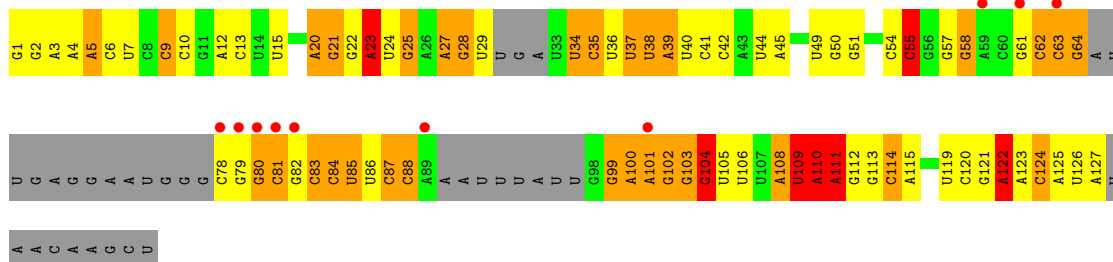
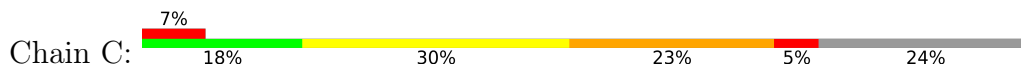


• Molecule 1: Selenomethionine (SeMet)-labeled Cas12c1 D969A mutant

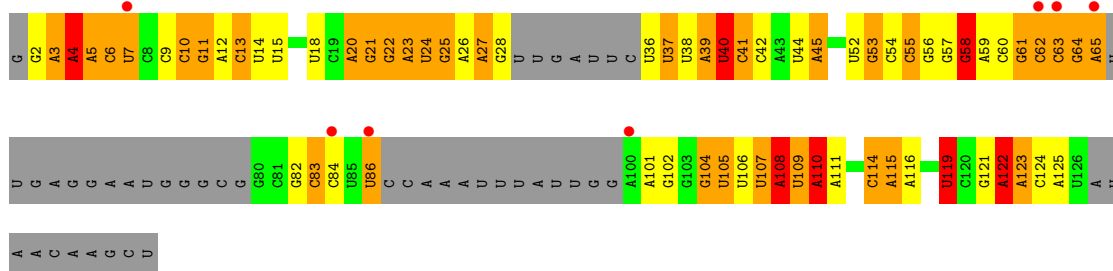




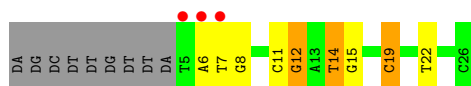
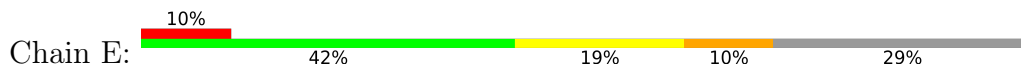
• Molecule 2: sgRNA



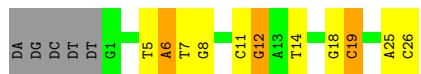
• Molecule 2: sgRNA



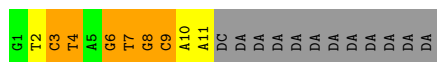
• Molecule 3: Target DNA strand




• Molecule 3: Target DNA strand

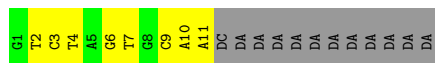


• Molecule 4: Non-target DNA strand



- Molecule 4: Non-target DNA strand

Chain H:  13% 35% 52%



4 Data and refinement statistics i

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	93.13Å 149.85Å 173.25Å 90.00° 91.59° 90.00°	Depositor
Resolution (Å)	49.73 – 3.20 49.68 – 3.20	Depositor EDS
% Data completeness (in resolution range)	79.4 (49.73-3.20) 79.4 (49.68-3.20)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.96 (at 3.19Å)	Xtrriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.256 , 0.292 0.256 , 0.292	Depositor DCC
R_{free} test set	2973 reflections (4.78%)	wwPDB-VP
Wilson B-factor (Å ²)	53.0	Xtrriage
Anisotropy	0.063	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.27 , 54.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.45$, $\langle L^2 \rangle = 0.27$	Xtrriage
Estimated twinning fraction	0.045 for h,-k,-l	Xtrriage
F_o, F_c correlation	0.82	EDS
Total number of atoms	24465	wwPDB-VP
Average B, all atoms (Å ²)	73.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.86% 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

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	1.20	8/9713 (0.1%)	0.96	31/13105 (0.2%)
1	B	1.09	4/9549 (0.0%)	0.95	34/12891 (0.3%)
2	C	0.81	11/2441 (0.5%)	0.89	6/3793 (0.2%)
2	D	0.70	6/2158 (0.3%)	0.96	9/3352 (0.3%)
3	E	0.68	0/504	1.63	9/776 (1.2%)
3	F	0.67	0/597	1.53	10/920 (1.1%)
4	G	1.25	4/254 (1.6%)	1.03	2/390 (0.5%)
4	H	0.59	0/254	0.91	0/390
All	All	1.06	33/25470 (0.1%)	0.99	101/35617 (0.3%)

The worst 5 of 33 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	337	PRO	N-CD	13.97	1.67	1.47
1	B	956	PRO	N-CD	10.04	1.61	1.47
1	A	515	PRO	N-CD	9.92	1.61	1.47
1	A	460	PRO	N-CD	-9.83	1.34	1.47
1	A	1190	PHE	C-N	9.49	1.55	1.34

The worst 5 of 101 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	E	19	DC	O5'-P-OP2	-20.37	86.26	110.70
3	F	14	DT	O5'-P-OP1	-18.06	89.03	110.70
3	F	19	DC	O5'-P-OP1	-17.79	89.36	110.70
3	E	14	DT	O5'-P-OP2	-16.98	90.33	110.70
3	E	12	DG	O5'-P-OP1	-16.18	91.14	105.70

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	9533	0	9379	994	0
1	B	9369	0	9165	1088	0
2	C	2188	0	1113	138	0
2	D	1934	0	984	123	0
3	E	451	0	251	10	0
3	F	534	0	297	7	0
4	G	227	0	125	15	0
4	H	227	0	125	6	0
5	A	1	0	0	0	0
5	B	1	0	0	0	0
All	All	24465	0	21439	2321	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 51.

The worst 5 of 2321 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:B:1159:TRP:CH2	1:B:1183:ALA:HB2	1.29	1.67
1:B:565:VAL:HB	1:B:708:ARG:CD	1.31	1.60
1:B:565:VAL:CB	1:B:708:ARG:HD3	1.34	1.52
1:A:444:VAL:HG21	1:A:471:PHE:CD2	1.43	1.49
1:A:1201:TYR:CE2	1:A:1209:PRO:HD2	1.48	1.47

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	1197/1310 (91%)	1130 (94%)	59 (5%)	8 (1%)	22	61
1	B	1177/1310 (90%)	1092 (93%)	68 (6%)	17 (1%)	11	46
All	All	2374/2620 (91%)	2222 (94%)	127 (5%)	25 (1%)	14	51

5 of 25 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	656	ARG
1	B	841	ARG
1	A	175	ASN
1	A	328	LYS
1	A	881	VAL

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	1018/1144 (89%)	750 (74%)	268 (26%)	0	2
1	B	993/1144 (87%)	730 (74%)	263 (26%)	0	2
All	All	2011/2288 (88%)	1480 (74%)	531 (26%)	0	2

5 of 531 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	953	LEU
1	B	1018	GLN
1	B	944	SER
1	B	1244	THR
1	A	961	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 55 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	126	GLN
1	B	528	GLN
1	B	1284	ASN
1	B	1072	GLN
1	B	154	ASN

5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	C	99/136 (72%)	51 (51%)	9 (9%)
2	D	87/136 (63%)	49 (56%)	5 (5%)
All	All	186/272 (68%)	100 (53%)	14 (7%)

5 of 100 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	C	2	G
2	C	5	A
2	C	7	U
2	C	13	C
2	C	15	U

5 of 14 RNA pucker outliers are listed below:

Mol	Chain	Res	Type
2	C	110	A
2	C	114	C
2	D	114	C
2	D	82	G
2	D	110	A

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 2 ligands modelled in this entry, 2 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 [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	1200/1310 (91%)	0.14	60 (5%) 28 16	12, 61, 132, 179	0
1	B	1182/1310 (90%)	0.23	66 (5%) 24 13	16, 72, 129, 161	0
2	C	103/136 (75%)	0.50	10 (9%) 7 4	10, 81, 167, 198	0
2	D	91/136 (66%)	0.66	7 (7%) 13 7	13, 97, 168, 199	0
3	E	22/31 (70%)	0.58	3 (13%) 3 2	9, 44, 145, 166	0
3	F	26/31 (83%)	-0.07	0 100 100	14, 63, 139, 152	0
4	G	11/23 (47%)	0.09	0 100 100	23, 41, 91, 122	0
4	H	11/23 (47%)	0.02	0 100 100	26, 38, 78, 132	0
All	All	2646/3000 (88%)	0.21	146 (5%) 25 14	9, 67, 139, 199	0

The worst 5 of 146 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	733	ASN	8.7
1	A	578	ASN	6.3
3	E	6	DA	5.8
1	A	1178	GLY	5.5
1	B	455	ASP	5.5

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](#)

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(\AA^2)	Q<0.9
5	ZN	B	1401	1/1	0.80	0.13	79,79,79,79	0
5	ZN	A	1401	1/1	0.96	0.10	61,61,61,61	0

6.5 Other polymers [i](#)

There are no such residues in this entry.