wwPDB X-ray Structure Validation Summary Report (i)

May 25, 2020 – 08:06 am BST

PDB ID	:	6SUQ
Title	:	Crystal Structure of TcdB2-TccC3-TEV
Authors	:	Roderer, D.; Schubert, E.; Sitsel, O.; Raunser, S.
Deposited on	:	2019-09-16
Resolution	:	3.70 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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

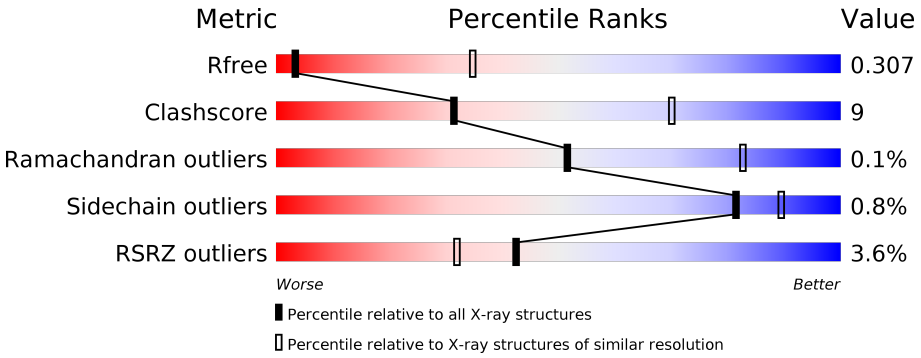
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 3.70 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	1049 (3.88-3.52)
Clashscore	141614	1027 (3.86-3.54)
Ramachandran outliers	138981	1069 (3.88-3.52)
Sidechain outliers	138945	1065 (3.88-3.52)
RSRZ outliers	127900	1578 (3.90-3.50)

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 <=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	2410	3% 69% 20% 11%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 17025 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 TcdB2,TccC3,Genome polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	2134	17025	10667	3024	3300	34	0	0	0

There are 21 discrepancies between the modelled and reference sequences:

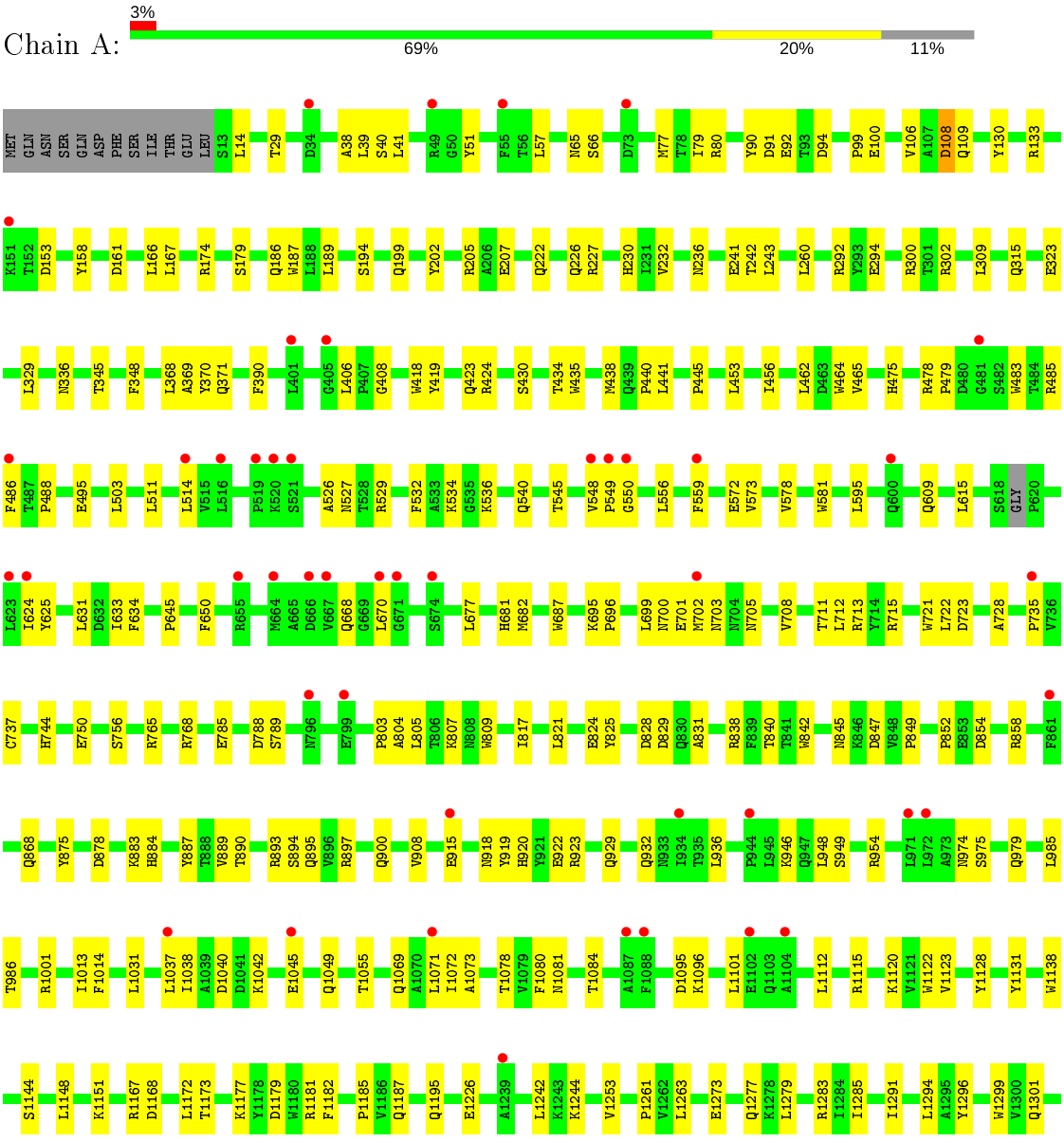
Chain	Residue	Modelled	Actual	Comment	Reference
A	1475	PRO	-	linker	UNP Q8GF99
A	1476	GLY	-	linker	UNP Q8GF99
A	1477	SER	-	linker	UNP Q8GF99
A	1478	ARG	-	linker	UNP Q8GF99
A	1479	PRO	-	linker	UNP Q8GF99
A	2160	GLU	-	linker	UNP Q8GF97
A	2161	PHE	-	linker	UNP Q8GF97
A	2380	VAL	SER	conflict	UNP P04517
A	2398	ARG	-	expression tag	UNP P04517
A	2399	ARG	-	expression tag	UNP P04517
A	2400	ARG	-	expression tag	UNP P04517
A	2401	ARG	-	expression tag	UNP P04517
A	2402	ARG	-	expression tag	UNP P04517
A	2403	LEU	-	expression tag	UNP P04517
A	2404	GLU	-	expression tag	UNP P04517
A	2405	HIS	-	expression tag	UNP P04517
A	2406	HIS	-	expression tag	UNP P04517
A	2407	HIS	-	expression tag	UNP P04517
A	2408	HIS	-	expression tag	UNP P04517
A	2409	HIS	-	expression tag	UNP P04517
A	2410	HIS	-	expression tag	UNP P04517

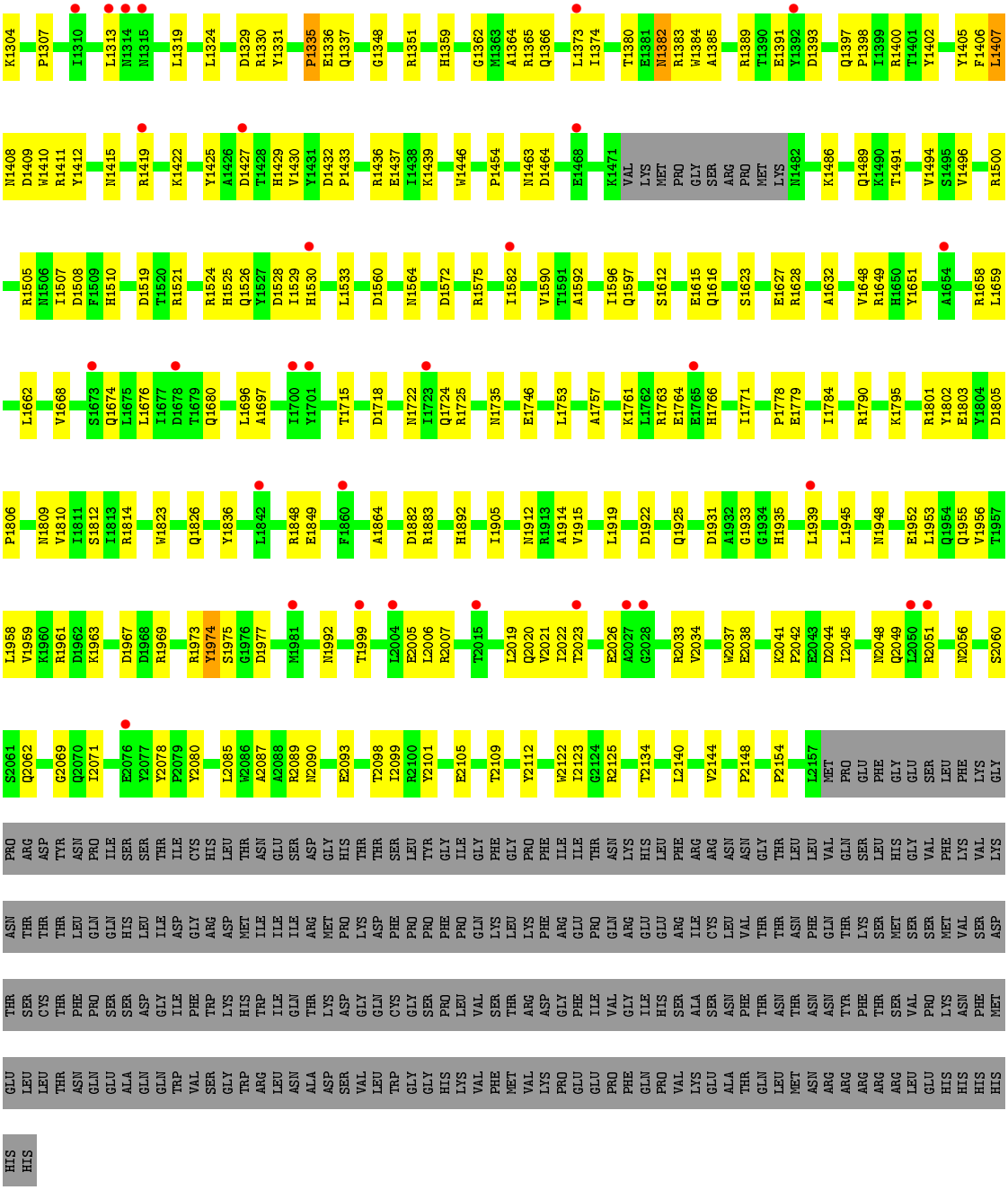
3 Residue-property plots (i)

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: TcdB2,TccC3,Genome polyprotein

Chain A:




K1304 P1307 I1310 L1313 N1314 N1315 L1319 L1324 D1329 R1330 Y1331 P1335 E1336 Q1337 G1348 R1351 H1359 G1362 M1363 A1364 R1365 Q1366 L1373 I1374 T1380 E1381 N1382 R1383 W1384 A1385 R1389 T1390 E1391 Y1392 D1393 Q1397 P1398 I1399 R1400 T1401 Y1402 Y1405 F1406 L1407
N1408 D1409 W1410 R1411 Y1412 N1415 R1419 K1422 Y1425 A1426 D1427 T1428 H1429 V1430 Y1431 D1432 P1433 R1436 E1437 I1438 K1439 W1446 P1454 N1463 D1464 E1468 K1471 VAL LYS MET PRO GLY SER ARG PRO MET LYS N1482 K1486 Q1489 K1490 T1491 V1494 S1495 V1496 R1500
R1505 W1506 I1507 D1508 F1509 H1510 D1519 T1520 R1521 R1524 H1525 Q1526 Y1527 D1528 I1529 H1530 L1533 D1560 N1564 D1572 R1575 I1582 V1590 T1591 A1592 I1596 Q1597 S1612 E1615 Q1616 S1623 E1627 R1628 A1632 V1648 R1649 H1650 Y1651 A1654 R1658 L1659
L1662 V1668 S1673 Q1674 L1675 L1676 I1677 D1678 T1679 Q1680 L1696 A1697 I1700 Y1701 T1715 D1718 N1722 I1723 Q1724 R1725 N1735 E1746 L1753 A1757 K1761 L1762 R1763 E1764 E1765 H1766 I1771 P1778 E1779 I1784 R1790 K1795 R1801 Y1802 E1803 Y1804 D1805
P1806 N1809 V1810 I1811 S1812 I1813 R1814 W1823 Q1826 Y1836 L1842 R1848 E1849 F1860 A1864 D1882 R1883 H1892 I1905 N1912 R1913 A1914 V1915 L1919 D1922 Q1925 D1931 A1932 G1933 G1934 H1935 L1939 L1945 N1948 E1952 L1953 Q1954 Q1955 V1956 T1957
L1958 V1959 K1960 R1961 D1962 K1963 D1967 D1968 R1969 R1973 Y1974 S1975 G1976 D1977 M1981 N1992 T1999 L2004 E2005 L2006 R2007 T2015 L2019 Q2020 V2021 I2022 T2023 E2026 A2027 G2028 R2033 V2034 W2037 E2038 K2041 P2042 E2043 D2044 I2045 N2048 Q2049 L2050 R2051 N2056 S2060
S2061 Q2062 G2069 Q2070 I2071 E2076 Y2077 Y2078 P2079 Y2080 L2085 W2086 A2087 A2088 R2089 N2090 E2093 T2098 I2099 R2100 Y2101 E2105 T2109 Y2112 W2122 I2123 G2124 R2125 T2134 L2140 V2144 P2148 P2154 L2157 MET PRO GLU PHE GLY GLU SER LEU PHE LYS VAL LYS
PRO ARG ASP TYR ASN PRO ILE SER SER THR ILE CYS HIS LEU THR ASN GLU SER ASP GLY HIS THR THR SER LEU TYR GLY ILE GLY PHE GLY PRO PHE ILE ILE THR ASN LYS HIS LEU PHE ARG ARG ASN ASN GLY THR THR LEU LEU VAL GLN SER LEU HIS GLY VAL PHE LYS VAL LYS
ASN THR THR THR LEU GLN GLN HIS LEU ILE ASP GLY ARG ASP MET PRO LYS ASP PHE PRO PRO PHE PRO GLN LYS LEU LYS PHE ARG GLU ILE PRO GLN ARG GLU GLU ARG ILE CYS LEU VAL THR THR ASN PHE GLN THR LYS SER MET SER SER MET VAL SER ASP
THR SER CYS THR PHE PRO SER SER ASP GLY ILE PHE TRP LYS HIS TRP ILE GLN THR LYS ASP GLY GLN CYS GLY SER PRO LEU VAL SER THR ARG ASP GLY PHE ILE VAL GLY ILE HIS SER ALA SER ASN PHE THR ASN THR ASN ASN TYR PHE THR SER VAL PRO LYS ASN PHE MET
GLU LEU LEU THR ASN GLN GLU ALA GLN GLN TRP VAL SER GLY TRP ARG LEU ASN ALA ASP SER VAL LEU TRP GLY GLY HIS LYS VAL PHE MET VAL LYS PRO GLU GLU PRO PHE GLN PRO VAL LYS GLU ALA THR GLN LEU MET ASN ARG ARG ARG ARG ARG LEU GLU HIS HIS HIS
HIS HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, α, β, γ	234.43Å 234.43Å 143.19Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	47.73 – 3.70 49.18 – 3.70	Depositor EDS
% Data completeness (in resolution range)	99.9 (47.73-3.70) 100.0 (49.18-3.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.01 (at 3.67Å)	Xtriage
Refinement program	PHENIX 1.16-3549	Depositor
R, R_{free}	0.269 , 0.302 0.286 , 0.307	Depositor DCC
R_{free} test set	2431 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	130.9	Xtriage
Anisotropy	0.362	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 96.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.44$, $< L^2 > = 0.26$	Xtriage
Estimated twinning fraction	0.196 for -h,-k,l	Xtriage
F_o,F_c correlation	0.87	EDS
Total number of atoms	17025	wwPDB-VP
Average B, all atoms (Å^2)	135.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.12% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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	0.26	0/17444	0.46	2/23779 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	1407	LEU	CA-CB-CG	6.46	130.16	115.30
1	A	1407	LEU	CB-CG-CD2	-6.00	100.80	111.00

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	17025	0	16395	302	0
All	All	17025	0	16395	302	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 9.

The worst 5 of 302 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:1330:ARG:HD3	1:A:1335:PRO:HB3	1.62	0.81

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:371:GLN:HB3	1:A:696:PRO:HA	1.66	0.76
1:A:534:LYS:HD3	1:A:536:LYS:HD3	1.71	0.72
1:A:668:GLN:HB3	1:A:670:LEU:HD13	1.73	0.70
1:A:91:ASP:OD1	1:A:92:GLU:N	2.23	0.70

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	2128/2410 (88%)	2033 (96%)	92 (4%)	3 (0%)	51 83

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	1335	PRO
1	A	2154	PRO
1	A	479	PRO

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	1842/2095 (88%)	1828 (99%)	14 (1%)	81 89

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1182	PHE
1	A	1382	ASN
1	A	1848	ARG
1	A	788	ASP
1	A	1575	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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, 95^{th} 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	2134/2410 (88%)	0.20	77 (3%) 42 32	94, 130, 186, 267	0

The worst 5 of 77 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	670	LEU	6.8
1	A	1315	ASN	6.2
1	A	548	VAL	5.2
1	A	2023	THR	4.8
1	A	1419	ARG	4.6

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.