wwPDB X-ray Structure Validation Summary Report (i)

May 13, 2020 – 10:10 am BST

PDB ID	:	5XXZ
Title	:	Crystal structure of a serine protease from Streptococcus species
Authors	:	Jobichen, C.; Sivaraman, J.
Deposited on	:	2017-07-05
Resolution	:	3.08 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:

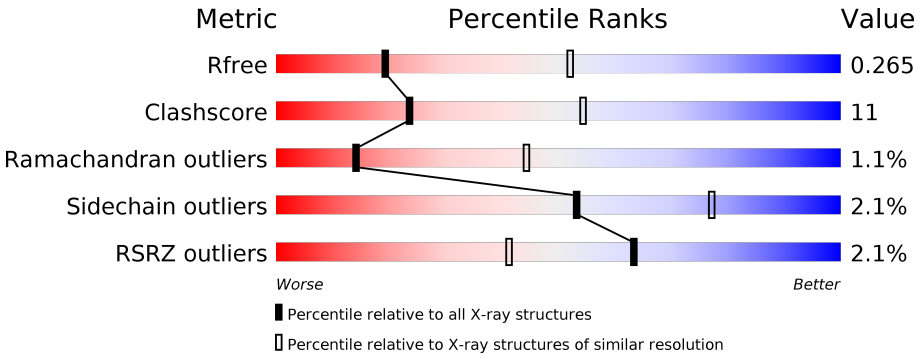
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 3.08 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	1447 (3.10-3.06)
Clashscore	141614	1546 (3.10-3.06)
Ramachandran outliers	138981	1487 (3.10-3.06)
Sidechain outliers	138945	1486 (3.10-3.06)
RSRZ outliers	127900	1416 (3.10-3.06)

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	1533	2% 64% 22% 13%
1	B	1533	% 64% 20% 15%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	CA	A	1705	-	-	X	-

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 19914 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.

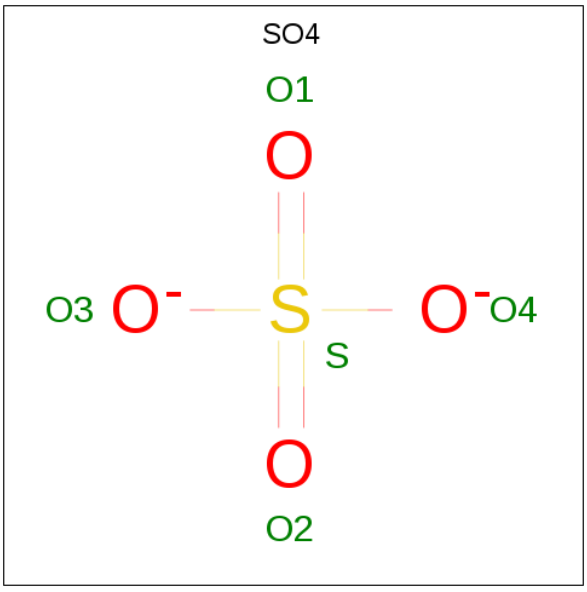
- Molecule 1 is a protein called Chemokine protease C.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1334	Total	C	N	O	Se	0	0	0
			10007	6297	1722	1964	24			
1	B	1310	Total	C	N	O	Se	0	0	0
			9890	6229	1701	1936	24			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	279	ALA	HIS	engineered mutation	UNP Q3HV58
A	617	ALA	SER	engineered mutation	UNP Q3HV58
B	279	ALA	HIS	engineered mutation	UNP Q3HV58
B	617	ALA	SER	engineered mutation	UNP Q3HV58

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total O S 5 4 1	0	0
2	B	1	Total O S 5 4 1	0	0

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

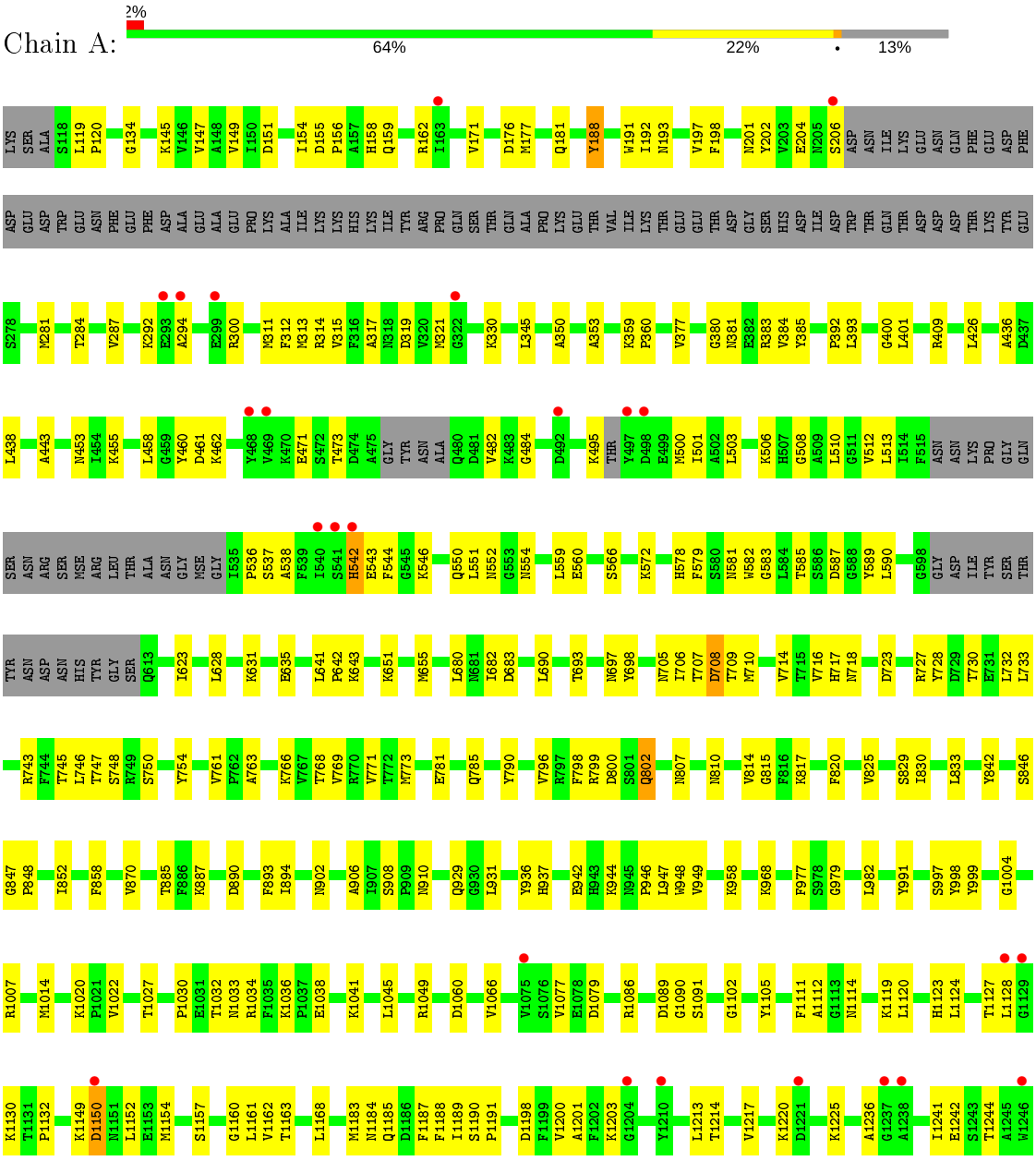
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	3	Total Ca 3 3	0	0
3	A	4	Total Ca 4 4	0	0

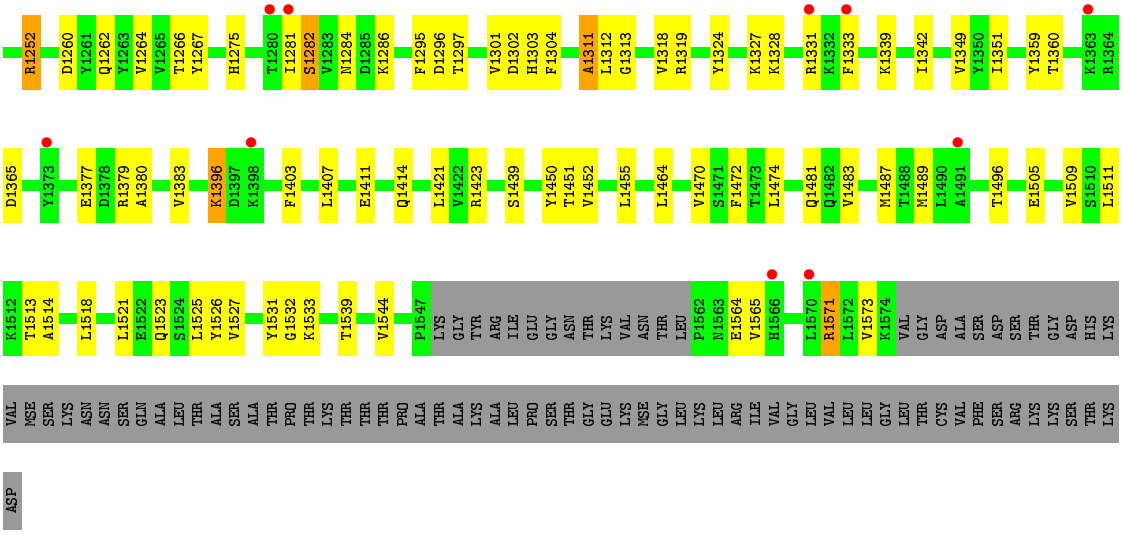
3 Residue-property plots (i)

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: Chemokine protease C

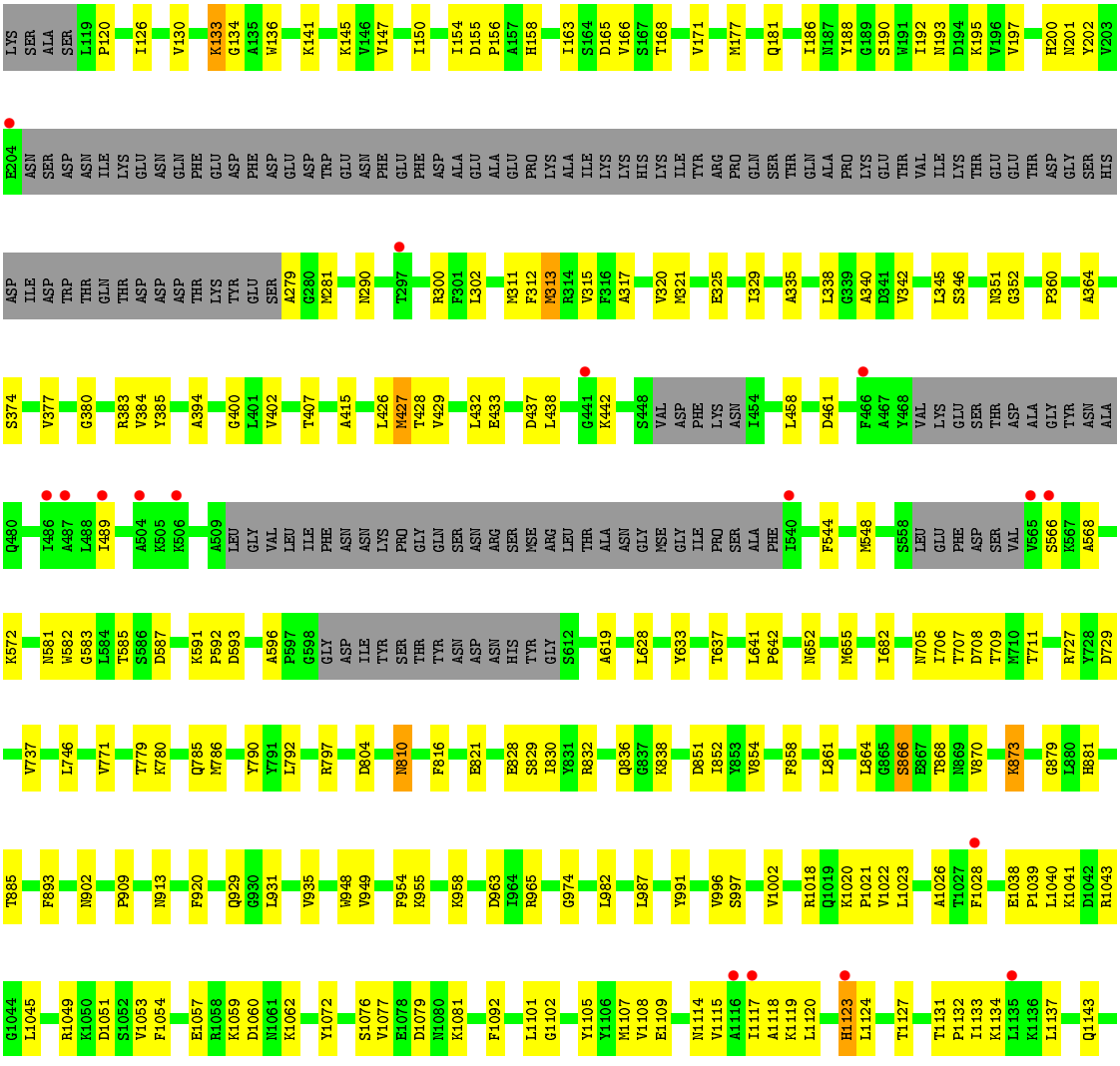
Chain A:

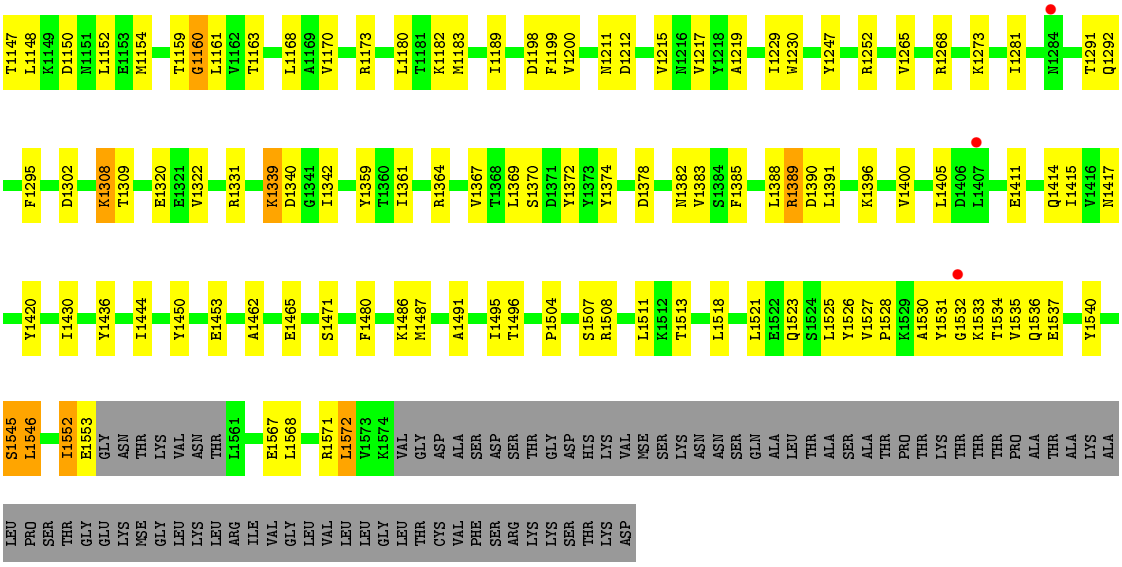




• Molecule 1: Chemokine protease C

Chain B:




T1147 L1148 K1149 D1150 N1151 L1152 E1153 M1154 T1159 G1160 L1161 V1162 T1163 L1168 A1169 V1170 R1173 L1180 T1181 K1182 M1183 I1189 D1198 F1199 V1200 N1211 D1212 V1215 N1216 V1217 Y1218 A1219 I1229 W1230 Y1247 R1252 V1265 R1268 K1273 I1281 N1284 T1291 Q1292
F1295 D1302 K1308 T1309 E1320 E1321 V1322 R1331 K1339 D1340 G1341 I1342 Y1359 T1360 I1361 R1364 V1367 T1368 L1369 S1370 D1371 Y1372 Y1373 Y1374 D1378 N1382 V1383 S1384 F1385 L1388 R1389 D1390 L1391 K1396 V1400 L1405 D1406 L1407 E1411 Q1414 I1415 V1416 N1417
Y1420 I1430 Y1436 I1444 Y1450 E1453 A1462 E1465 S1471 F1480 K1486 M1487 A1491 I1495 T1496 P1504 S1507 R1508 L1511 K1512 T1513 L1518 L1521 E1522 Q1523 S1524 L1525 Y1526 V1527 P1528 K1529 A1530 Y1531 G1532 K1533 T1534 V1535 Q1536 E1537 Y1540
S1545 L1546 I1552 E1553 GLY ASN THR LYS VAL ASN THR L1561 E1567 L1568 R1571 L1572 V1573 K1574 VAL GLY ASP ALA SER ASP SER THR GLY ASP HIS LYS VAL MSE SER LYS ASN ASN SER GLN ALA LEU THR ALA SER ALA THR PRO THR LYS THR THR THR PRO ALA THR ALA LYS ALA
LEU PRO SER THR GLY GLU LYS MSE GLY LEU LYS LEU ARG ILE VAL GLY LEU VAL LEU LEU GLY LEU THR CYS VAL PHE SER ARG LYS LYS SER THR LYS ASP

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	182.13Å 132.85Å 151.99Å 90.00° 100.69° 90.00°	Depositor
Resolution (Å)	19.98 – 3.08 49.78 – 3.09	Depositor EDS
% Data completeness (in resolution range)	98.7 (19.98-3.08) 93.0 (49.78-3.09)	Depositor EDS
R_{merge}	0.19	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.07 (at 3.07Å)	Xtriage
Refinement program	PHENIX (dev_2733: ???)	Depositor
R, R_{free}	0.205 , 0.263 0.211 , 0.265	Depositor DCC
R_{free} test set	1980 reflections (3.07%)	wwPDB-VP
Wilson B-factor (Å^2)	51.1	Xtriage
Anisotropy	0.458	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 25.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.87	EDS
Total number of atoms	19914	wwPDB-VP
Average B, all atoms (Å^2)	61.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.18% 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

Bond lengths and bond angles in the following residue types are not validated in this section: CA, SO4

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	0.60	0/10170	0.76	1/13731 (0.0%)
1	B	0.60	0/10051	0.75	1/13571 (0.0%)
All	All	0.60	0/20221	0.75	2/27302 (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	B	1198	ASP	CB-CG-OD2	-5.81	113.07	118.30
1	A	1198	ASP	CB-CG-OD1	5.09	122.88	118.30

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	10007	0	9601	224	0
1	B	9890	0	9560	215	0
2	A	5	0	0	0	0
2	B	5	0	0	0	0
3	A	4	0	0	2	0
3	B	3	0	0	0	0
All	All	19914	0	19161	439	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 11.

The worst 5 of 439 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:433:GLU:CA	1:B:437:ASP:CB	2.10	1.30
1:A:1496:THR:HA	1:A:1532:GLY:HA3	1.36	1.05
1:A:1190:SER:OG	3:A:1705:CA:CA	1.48	0.90
1:B:1496:THR:HA	1:B:1532:GLY:HA3	1.51	0.90
1:B:1147:THR:HG22	1:B:1148:LEU:H	1.40	0.86

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	1320/1533 (86%)	1182 (90%)	117 (9%)	21 (2%)	9 35
1	B	1294/1533 (84%)	1165 (90%)	121 (9%)	8 (1%)	25 57
All	All	2614/3066 (85%)	2347 (90%)	238 (9%)	29 (1%)	14 44

5 of 29 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	537	SER
1	A	538	ALA
1	A	554	ASN
1	A	1038	GLU
1	A	1184	ASN

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	1036/1271 (82%)	1015 (98%)	21 (2%)	55 78
1	B	1033/1271 (81%)	1011 (98%)	22 (2%)	53 77
All	All	2069/2542 (81%)	2026 (98%)	43 (2%)	53 77

5 of 43 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1439	SER
1	B	427	MSE
1	B	1389	ARG
1	A	1571	ARG
1	B	133	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	200	HIS

5.3.3 RNA

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates

There are no carbohydrates in this entry.

5.6 Ligand geometry

Of 9 ligands modelled in this entry, 7 are monoatomic - leaving 2 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	SO4	B	1701	-	4,4,4	0.28	0	6,6,6	0.59	0
2	SO4	A	1701	-	4,4,4	0.19	0	6,6,6	0.42	0

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

There are no such residues in this entry.

5.8 Polymer linkage issues

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	1309/1533 (85%)	0.02	34 (2%)	56	31	32, 57, 117, 196	0
1	B	1285/1533 (83%)	-0.09	20 (1%)	72	51	30, 54, 111, 204	0
All	All	2594/3066 (84%)	-0.04	54 (2%)	63	41	30, 56, 115, 204	0

The worst 5 of 54 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1129	GLY	5.5
1	A	498	ASP	4.9
1	A	492	ASP	4.5
1	A	1128	LEU	4.5
1	B	506	LYS	4.1

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

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, 95^{th} 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
3	CA	A	1704	1/1	0.83	0.12	53,53,53,53	0
2	SO4	B	1701	5/5	0.86	0.27	98,98,99,99	0
2	SO4	A	1701	5/5	0.87	0.30	115,115,116,116	0
3	CA	A	1702	1/1	0.91	0.35	74,74,74,74	0
3	CA	B	1703	1/1	0.94	0.06	48,48,48,48	0
3	CA	B	1704	1/1	0.96	0.10	43,43,43,43	0
3	CA	B	1702	1/1	0.98	0.04	35,35,35,35	0
3	CA	A	1703	1/1	0.98	0.14	37,37,37,37	0
3	CA	A	1705	1/1	0.98	0.17	30,30,30,30	0

6.5 Other polymers (i)

There are no such residues in this entry.