



wwPDB X-ray Structure Validation Summary Report

Nov 13, 2023 – 07:58 PM JST

PDB ID : 5XYA
Title : Crystal structure of a serine protease from Streptococcus species
Authors : Jobichen, C.; Sivaraman, J.
Deposited on : 2017-07-06
Resolution : 3.00 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 : 2.36
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) : 2.36

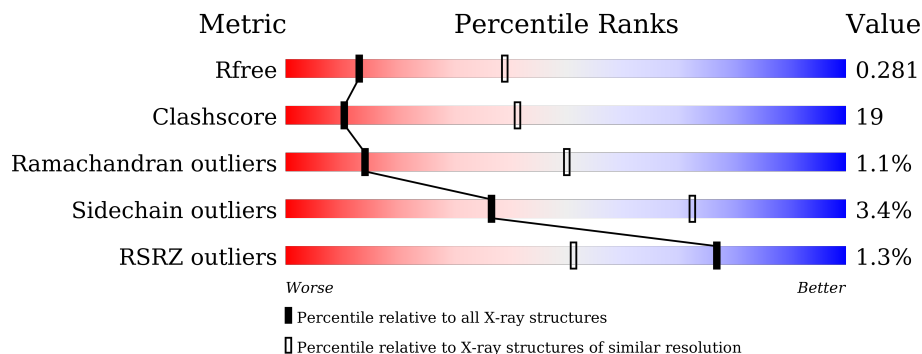
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.00 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	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)

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	1530	

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	SO4	A	1704	-	-	X	-

2 Entry composition [i](#)

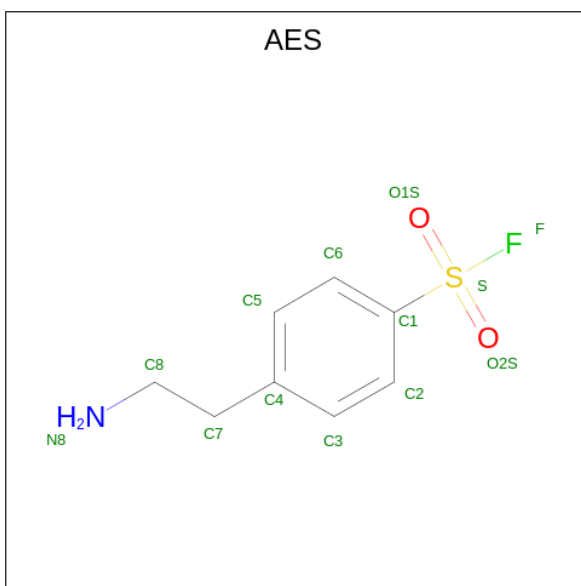
There are 4 unique types of molecules in this entry. The entry contains 10300 atoms, of which 0 are hydrogens and 0 are deuteriums.

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
			Total	C	N	O	Se			
1	A	1358	10269	6462	1764	2019	24	0	0	0

- Molecule 2 is 4-(2-AMINOETHYL)BENZENESULFONYL FLUORIDE (three-letter code: AES) (formula: C₈H₁₀FNO₂S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	S		
2	A	1	12	8	1	2	1	0	0

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



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

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

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

LEU	V1048	E1153	M1257	D1365	K1486	K1574	THR
THR	R1049	M1154	V1265	G1366	M1487	VAL	LEU
ALA	F1054	V1162	T1266	V1367	T1368	GLY	LEU
ALA	Y1055	Y1267	L1369	T1368	L1369	ASP	THR
THR	L1056	A1166	Y1267	L1369	T1496	ALA	VAL
PRO	D1060	H1275	H1275	S1370	F1499	ALA	PHE
THR	P1063	Q1276	Q1276	D1371	D1500	ASP	SER
LYS	Y1064	K1277	Q1277	Y1372	H1501	THR	ARG
THR	T1065	S1178	Q1278	Y1373	L1502	LYS	LYS
THR	V1066	Q1179	Q1278	Y1374	L1503	THR	LYS
ALA	T1067	L1180	T1280	E1377	L1521	THR	SER
THR	I1068	F1187	I1281	A1380	E1522	PRO	LYS
ALA	V1075	F1188	S1282	V1383	Q1523	ALA	ASP
ALA	S1076	F1189	G1293	V1383	Y1526	ALA	THR
LEU	V1077	I1189	R1294	L1388	Y1526	LYS	LEU
PRO	M1080	S1190	M1299	L1388	Y1531	ALA	LEU
SER	K1081	P1191	G1300	K1396	G1532	LEU	LEU
THR	R1086	M1192	D1302	D1397	K1533	THR	LEU
GLY	K1197	E1193	H1303	V1400	V1544	GLY	LEU
GLU	F1200	D1194	H1303	F1403	L1546	LEU	LEU
LYS	F1092	K1197	D1307	F1403	P1547	MSE	LEU
GLY	D1097	V1200	S1314	D1406	K1548	GLY	LEU
LYS	G1102	K1203	E1320	P1410	G1549	LEU	LEU
LEU	Y1105	G1204	E1320	P1410	Y1550	LEU	LEU
LEU	Y1106	L1205	F1323	I1415	R1551	LEU	LEU
LEU	M1107	L1205	Y1324	Y1416	I1552	LEU	LEU
LEU	V1108	M1208	L1325	M1417	E1553	LEU	LEU
LEU	D1110	Y1209	R1331	R1423	E1564	GLY	LEU
LEU	F1111	Y1210	K1332	D1424	V1565	LEU	LEU
LEU	A1112	L1214	E1337	A1425	R1571	LEU	LEU
LEU	G1113	Y1218	T1342	Y1435	K1574	GLY	LEU
CYS	N1114	H1223	T1343	G1440	VAL	LEU	LEU
VAL	V1115	H1223	V1344	T1343	THR	THR	THR
PHE	A1116	Q1224	V1344	K1449	ASP	THR	THR
SER	I1117	K1225	M1347	Y1450	ASP	THR	THR
SER	A1118	Q1233	K1348	T1451	THR	THR	THR
LYS	K1119	A1236	V1349	V1452	THR	THR	THR
LYS	L1120	A1236	I1351	L1455	ASP	THR	THR
LYS	L1124	S1239	P1352	L1464	ASP	THR	THR
LYS	P1125	A1240	K1353	E1466	VAL	THR	THR
LYS	Q1126	L1241	D1356	S1466	MSE	THR	THR
LYS	T1127	E1242	G1357	S1466	MSE	THR	THR
LYS	L1128	V1246	S1358	S1475	SER	THR	THR
LYS	G1129	Y1247	Y1359	A1476	SER	THR	THR
LYS	K1130	G1248	T1360	D1477	LYS	THR	THR
LYS	T1131	R1252	I1361	F1480	ASN	THR	THR
LYS	Y1142	V1256	K1362	V1483	ASN	THR	THR
LYS	L1152	V1256	R1363	V1483	GLN	THR	THR
LYS	L1152	V1256	R1364	V1483	ALA	THR	THR

4 Data and refinement statistics

Property	Value	Source
Space group	P 62 2 2	Depositor
Cell constants a, b, c, α , β , γ	190.53Å 190.53Å 248.67Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	19.90 – 3.00 49.84 – 2.85	Depositor EDS
% Data completeness (in resolution range)	99.4 (19.90-3.00) 88.1 (49.84-2.85)	Depositor EDS
R_{merge}	0.17	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.72 (at 2.86Å)	Xtrriage
Refinement program	PHENIX (dev_2733: ???)	Depositor
R, R_{free}	0.223 , 0.278 0.227 , 0.281	Depositor DCC
R_{free} test set	1990 reflections (3.60%)	wwPDB-VP
Wilson B-factor (Å ²)	37.2	Xtrriage
Anisotropy	0.047	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.26 , 30.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.40$, $\langle L^2 \rangle = 0.22$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.85	EDS
Total number of atoms	10300	wwPDB-VP
Average B, all atoms (Å ²)	42.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.61% 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: SO4, AES, CA

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	0.62	0/10444	0.78	4/14117 (0.0%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	1299	ASN	CB-CA-C	6.47	123.35	110.40
1	A	536	PRO	N-CA-CB	6.29	110.85	103.30
1	A	493	PRO	N-CA-CB	6.11	110.63	103.30
1	A	610	TYR	CB-CA-C	5.53	121.45	110.40

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	10269	0	9868	375	0
2	A	12	0	10	2	0
3	A	15	0	0	2	0
4	A	4	0	0	0	0
All	All	10300	0	9878	375	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 19.

The worst 5 of 375 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:188:TYR:OH	1:A:330:LYS:HE3	1.32	1.25
1:A:401:LEU:O	1:A:581:ASN:ND2	1.82	1.12
1:A:412:THR:HG22	1:A:651:LYS:HD3	1.40	1.04
1:A:910:ASN:ND2	1:A:912:ASP:OD1	1.94	1.00
1:A:188:TYR:OH	1:A:330:LYS:CE	2.13	0.95

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	1350/1530 (88%)	1159 (86%)	176 (13%)	15 (1%)	14 50

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	536	PRO
1	A	581	ASN
1	A	613	GLN
1	A	582	TRP
1	A	564	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	1073/1271 (84%)	1036 (97%)	37 (3%)	37 72

5 of 37 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1208	ASN
1	A	1477	ASP
1	A	1239	SER
1	A	1314	SER
1	A	591	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	418	ASN
1	A	581	ASN
1	A	881	HIS
1	A	810	ASN
1	A	309	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 8 ligands modelled in this entry, 4 are monoatomic - leaving 4 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
3	SO4	A	1704	-	4,4,4	0.54	0	6,6,6	0.82	0
3	SO4	A	1702	-	4,4,4	0.65	0	6,6,6	0.58	0
3	SO4	A	1703	-	4,4,4	0.41	0	6,6,6	0.56	0
2	AES	A	1701	1	8,12,13	0.69	0	13,15,18	1.67	3 (23%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	AES	A	1701	1	-	3/7/7/9	0/1/1/1

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	1701	AES	C5-C6-C1	3.47	122.49	119.40
2	A	1701	AES	C2-C1-C6	-3.46	117.72	121.59
2	A	1701	AES	O2S-S-C1	2.18	109.55	104.58

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	1701	AES	C2-C1-S-O2S
2	A	1701	AES	C6-C1-S-O2S
2	A	1701	AES	C4-C7-C8-N8

There are no ring outliers.

2 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	1704	SO4	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1701	AES	2	0

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	1333/1530 (87%)	-0.54	17 (1%) 77 51	6, 39, 88, 117	0

The worst 5 of 17 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1126	GLN	3.3
1	A	1129	GLY	3.1
1	A	465	GLN	3.0
1	A	538	ALA	3.0
1	A	1125	PRO	2.9

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	CA	A	1708	1/1	0.80	0.34	57,57,57,57	0
2	AES	A	1701	12/13	0.91	0.24	59,81,87,87	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
3	SO4	A	1702	5/5	0.92	0.26	28,29,60,83	0
3	SO4	A	1704	5/5	0.94	0.34	35,36,48,81	0
4	CA	A	1707	1/1	0.96	0.14	21,21,21,21	0
3	SO4	A	1703	5/5	0.96	0.17	37,42,57,73	0
4	CA	A	1705	1/1	0.98	0.10	18,18,18,18	0
4	CA	A	1706	1/1	0.98	0.10	24,24,24,24	0

6.5 Other polymers [\(i\)](#)

There are no such residues in this entry.