



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 21, 2020 – 09:47 AM BST

PDB ID : 6OQ5
Title : Structure of the full-length Clostridium difficile toxin B in complex with 3 VHHs
Authors : Chen, P.; Lam, K.; Jin, R.
Deposited on : 2019-04-25
Resolution : 3.87 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 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 : 2.13.1
buster-report : 1.1.7 (2018)
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) : 2.13.1

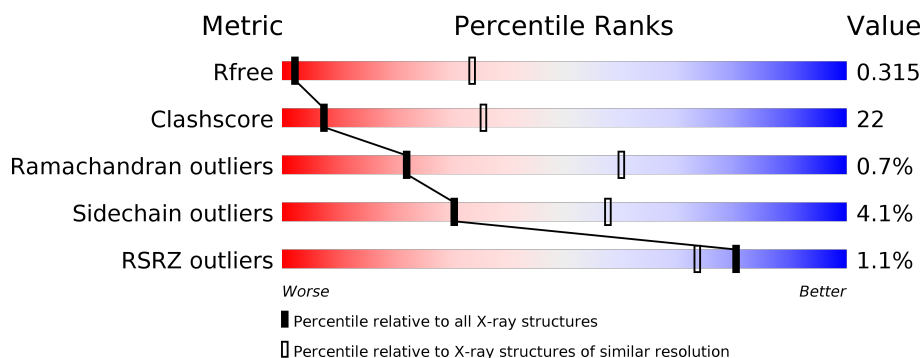
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.87 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	1026 (4.12-3.64)
Clashscore	141614	1045 (4.10-3.66)
Ramachandran outliers	138981	1008 (4.10-3.66)
Sidechain outliers	138945	1001 (4.10-3.66)
RSRZ outliers	127900	1213 (4.16-3.60)

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 $\leq 5\%$. 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	2373	<div> <div style="width: 9%;"></div> <div style="width: 62%;"></div> <div style="width: 35%;"></div> <div style="width: 1%;"></div> </div>
2	D	153	<div> <div style="width: 9%;"></div> <div style="width: 53%;"></div> <div style="width: 25%;"></div> <div style="width: 18%;"></div> </div>
3	E	137	<div> <div style="width: 9%;"></div> <div style="width: 56%;"></div> <div style="width: 21%;"></div> <div style="width: 22%;"></div> </div>
4	F	142	<div> <div style="width: 54%;"></div> <div style="width: 26%;"></div> <div style="width: 18%;"></div> </div>

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 21503 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 Toxin B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	2346	Total	C	N	O	S	0	0	0
			18837	12009	2961	3820	47			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	2368	HIS	-	expression tag	UNP M4NKV9
A	2369	HIS	-	expression tag	UNP M4NKV9
A	2370	HIS	-	expression tag	UNP M4NKV9
A	2371	HIS	-	expression tag	UNP M4NKV9
A	2372	HIS	-	expression tag	UNP M4NKV9
A	2373	HIS	-	expression tag	UNP M4NKV9

- Molecule 2 is a protein called 5D.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	126	Total	C	N	O	S	0	0	0
			986	617	179	187	3			

- Molecule 3 is a protein called E3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	E	107	Total	C	N	O	S	0	0	0
			802	498	141	159	4			

- Molecule 4 is a protein called 7F.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	F	116	Total	C	N	O	S	0	0	0
			876	544	154	172	6			

- Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn) (labeled as "Ligand of Inter-

est" by author).

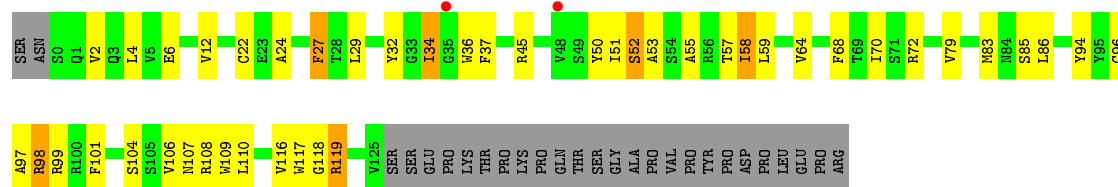
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	1	Total	Zn	0	0
			1	1		

- Molecule 6 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

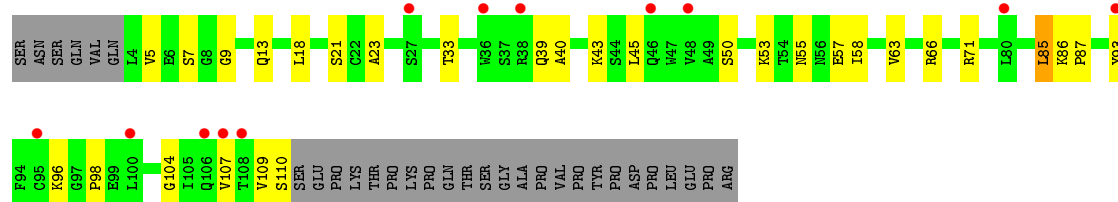
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	1	Total	Mg	0	0
			1	1		

L2329	E2228	T2117	Y2010	L1930	L1836	H1753	I1657	G1530	S1441	I1309	Y1204	M1109	L1026
D2330	T2229	M2118	F2018	L1930	I1837	S1754	N1661	Y1531	M1442	R1310	D1205	M1110	E1028
F2336	R2231	Q2119	Y2019	D1933	I1838	T1755	Y1662	Y1532	H1443	E1311	V1206	E1111	E1027
A2342	A2232	G2121	F2020	Y1938	I1840	E1757	D1663	L1545	Q1445	K1312	L1209	L1112	E1028
V2347	R2234	F2122	A2021	F1939	S1842	K1760	D1665	V1563	I1448	S1314	Q1210	R1115	P1031
F2356	L2235	T2125	E2022	E1940	L1843	F1768	D1666	V1563	I1448	Y1315	K1211	K1120	I1E
A2361	L2237	V2129	G2024	D1941	Y1844	Q1763	S1667	I1566	T1451	S1320	L1216	F1130	ALA
Q2362	L2238	F2130	E2025	N1942	Y1845	V1764	G1668	T1566	G1452	G1321	L1220	I1E	THR
Q2363	L2239	Y2131	Q2027	R1944	K1847	R1767	I1670	M1571	F1453	Y1331	L1223	I1E	ASP
L2363	E2241	F2132	I2028	Y1947	P1849	F1768	S1672	M1577	I1461	D1343	E1134	GLY	ASP
I2365	K2242	I2137	F2031	E1948	V1850	N1784	T1677	M1577	P1482	Y1344	P1227	L1140	VAL
H2369	K2243	I2138	T2033	N1949	N1851	F1785	I1674	S1579	S1464	W1345	N1228	D1141	SER
HIS	Y2245	I2145	E2034	E1951	L1853	K1788	S1678	K1592	E1472	I1346	R1229	K1143	LEU
HIS	F2246	Y2171	D2035	G1954	T1854	D1789	K1679	F1605	F1475	V1349	A1232	V1144	ALA
HIS	M2252	D2153	G2036	E1955	T1855	D1790	K1680	F1605	F1475	D1350	W1233	Q1148	ALA
	T2254	I2157	K2038	M1956	F1857	V1793	Y1681	M1610	T1476	M1351	W1237	D1149	LVS
	G2255	Y2171	E2042	Y1958	T1859	I1796	S1695	I1612	T1480	T1357	P1239	L1151	L1049
	L2256	Y2171	D2046	S1960	I1860	L1798	H1697	I1613	T1484	I1358	G1240	V1152	S1050
	M2262	M2181	L2047	G1964	D1863	S1799	I1698	T1616	F1486	K1362	R1242	E1155	E1051
	Y2265	I2182	E2051	F1967	Y1865	T1801	D1701	M1619	F1486	I1363	W1238	F1158	S1053
	F2266	Y2183	G2052	F1972	I1867	P1802	E1488	G1620	E1487	K1364	D1150	P1055	D1054
	M2267	V2187	E2053	L1970	N1868	Y1804	I1703	Q1621	L1489	K1365	V1164	L1056	L1056
	M2272	I2192	A2076	N1971	P1869	Y1805	T1706	F1622	P1490	I1372	K1250	L1165	E1060
	Y2285	V2193	W2080	I1973	I1870	E1806	V1708	F1624	D1491	L1373	L1166	L1165	I1061
	F2286	Y2200	S2087	G1974	S1876	D1807	Y1715	G1626	L1494	T1375	R1254	I1170	K1064
	Y2291	Y2201	K2088	D1975	I1881	G1811	I1719	D1627	I1495	I1378	I1255	R1171	K1064
	M2292	F2202	Y2090	K1977	I1882	I1812	V1720	E1628	S1496	E1379	E1262	R1172	I1087
	G2295	T2208	A2098	Y1979	N1886	L1814	N1724	M1631	Y1499	K1382	W1265	E1174	V1070
	M2298	E2209	Y2099	F1980	F1839	G1815	Y1725	Q1633	K1504	M1391	F1268	S1177	T1073
	F2303	G2211	L2100	N1981	I1894	V1817	N1727	P1634	P1505	E1397	A1272	H1179	T1074
	K2304	I2212	L2102	Q1987	L1895	S1818	I1726	Y1635	K1504	M1397	A1272	T1180	T1074
	Y2305	Y2214	L2104	Q1987	Q1896	L1819	N1732	F1639	Y1510	V1403	L1275	T1185	I1080
	F2306	D2215	I2105	I1993	S1901	K1823	N1733	E1643	N1512	T1406	I1276	L1186	L1084
	Q2309	M2216	D2107	N1994	F1906	Y1824	N1734	Y1646	N1514	F1187	K1280	F1187	S1091
	M2310	E2219	G2108	D1995	K1907	I1826	N1735	T1647	K1515	F1189	Y1283	F1189	I1092
	T2311	S2220	Q2109	N1996	Y1908	N1827	I1738	Y1648	D1516	A1415	N1287	P1192	L1093
	N2315	D2221	Y2110	H1998	F1908	F1829	I1739	V1650	T1521	I1416	I1417	S1193	L1094
	F2316	K2222	Y2111	Y1999	L1915	G1830	Y1741	M1652	K1522	E1418	S1298	T1195	V1095
	E2317	Y2224	F2112	F2000	M1918	M1831	W1742	M1652	K1522	V1419	P1302	Y1196	I1100
	T2321	F2225	D2114	V2005	A1923	M1832	W1743	R1653	Y1525	L1422	E1307	R1197	I1104
		D2226	D2115	M2006	I1924	S1834	I1751	Q1654	N1526	E1307	E1307	E1198	L1107
		F2227	G2116			G1835	L1752	M1656	T1529	M1440	Y1308	T1203	V1108

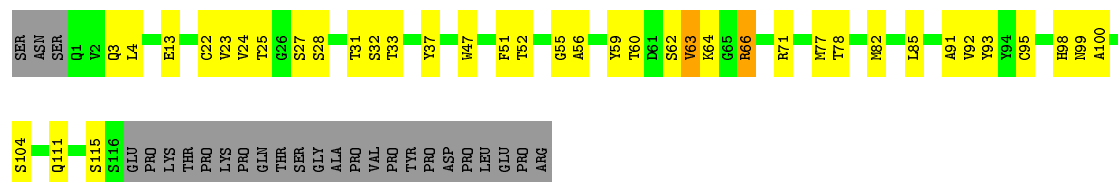
• Molecule 2: 5D



• Molecule 3: E3



• Molecule 4: 7F



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	149.62Å 168.56Å 179.92Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.91 – 3.87 48.87 – 3.87	Depositor EDS
% Data completeness (in resolution range)	99.2 (48.91-3.87) 99.3 (48.87-3.87)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.37 (at 3.88Å)	Xtriage
Refinement program	REFMAC 5.8.0232	Depositor
R, R_{free}	0.263 , 0.315 0.263 , 0.315	Depositor DCC
R_{free} test set	2203 reflections (5.13%)	wwPDB-VP
Wilson B-factor (Å ²)	135.9	Xtriage
Anisotropy	0.169	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.24 , 78.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.44$, $\langle L^2 \rangle = 0.26$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	21503	wwPDB-VP
Average B, all atoms (Å ²)	169.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.86% 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, MG

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	0/19208	0.64	0/26012
2	D	0.29	0/1006	0.65	0/1360
3	E	0.29	0/814	0.61	0/1098
4	F	0.32	0/893	0.65	0/1206
All	All	0.27	0/21921	0.64	0/29676

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	18837	0	18156	803	0
2	D	986	0	959	87	0
3	E	802	0	797	30	0
4	F	876	0	856	40	0
5	A	1	0	0	0	0
6	A	1	0	0	0	0
All	All	21503	0	20768	944	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 22.

The worst 5 of 944 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:2107:ASP:CB	1:A:2137:ILE:HG22	1.73	1.19
1:A:1847:LYS:HG3	1:A:1848:PRO:HD2	1.27	1.16
1:A:2025:GLU:O	1:A:2026:MET:HG2	1.44	1.14
1:A:120:ASP:HB3	1:A:357:LYS:HD2	1.18	1.14
1:A:625:THR:HB	1:A:628:GLU:HG2	1.25	1.14

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	2340/2373 (99%)	2001 (86%)	319 (14%)	20 (1%)	17	54
2	D	124/153 (81%)	112 (90%)	12 (10%)	0	100	100
3	E	105/137 (77%)	94 (90%)	11 (10%)	0	100	100
4	F	114/142 (80%)	98 (86%)	16 (14%)	0	100	100
All	All	2683/2805 (96%)	2305 (86%)	358 (13%)	20 (1%)	22	60

5 of 20 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	164	PRO
1	A	306	SER
1	A	307	VAL
1	A	574	SER
1	A	576	ARG

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	2113/2142 (99%)	2030 (96%)	83 (4%)	32	59
2	D	102/127 (80%)	93 (91%)	9 (9%)	10	37
3	E	89/117 (76%)	86 (97%)	3 (3%)	37	62
4	F	95/119 (80%)	92 (97%)	3 (3%)	39	63
All	All	2399/2505 (96%)	2301 (96%)	98 (4%)	30	58

5 of 98 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1680	LYS
1	A	1855	THR
2	D	106	VAL
1	A	1706	THR
1	A	1816	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 43 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	803	GLN
1	A	1445	GLN
1	A	2027	GLN
1	A	858	ASN
1	A	875	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 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	2346/2373 (98%)	-0.33	16 (0%) 87 82	80, 160, 233, 324	0
2	D	126/153 (82%)	-0.19	2 (1%) 72 63	147, 200, 255, 290	0
3	E	107/137 (78%)	0.62	12 (11%) 5 5	213, 260, 307, 341	0
4	F	116/142 (81%)	-0.42	0 100 100	117, 166, 214, 240	0
All	All	2695/2805 (96%)	-0.29	30 (1%) 80 73	80, 165, 249, 341	0

The worst 5 of 30 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	E	108	THR	4.1
1	A	1810	ILE	3.7
3	E	107	VAL	3.6
1	A	317	GLN	3.5
3	E	93	TYR	3.4

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](#)

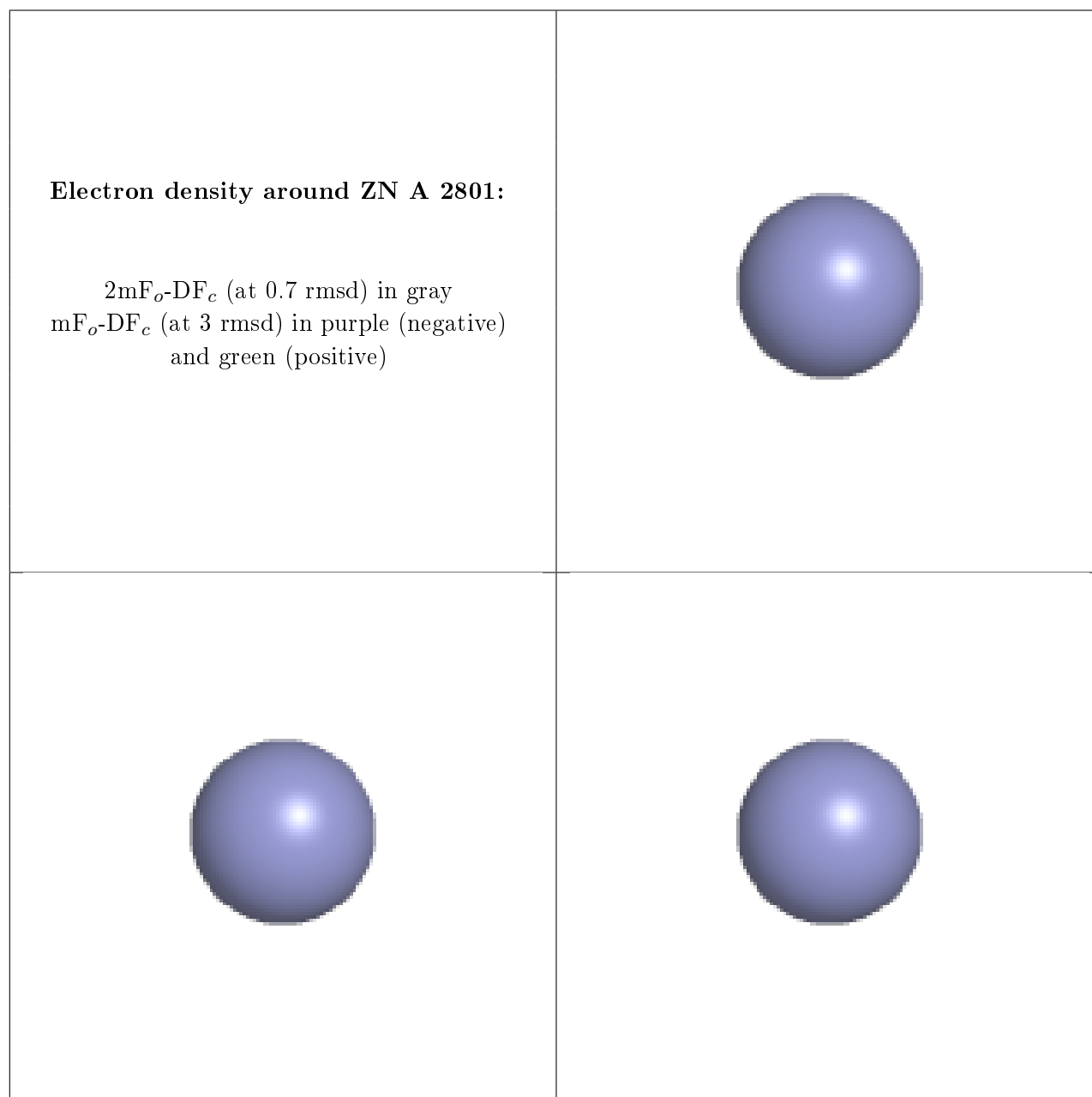
There are no monosaccharides in this entry.

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
6	MG	A	2802	1/1	0.80	0.47	119,119,119,119	0
5	ZN	A	2801	1/1	0.98	0.20	120,120,120,120	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.



6.5 Other polymers [i](#)

There are no such residues in this entry.