



wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 26, 2024 – 05:16 AM EDT

PDB ID : 6TS8
Title : Chaetomium thermophilum UDP-Glucose Glucosyl Transferase (UGGT) double cysteine mutant G177C/A786C.
Authors : Roversi, P.; Zitzmann, N.; Ibba, R.; Hensen, M.; Chandran, A.
Deposited on : 2019-12-20
Resolution : 4.60 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
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
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)	:	2.37.1

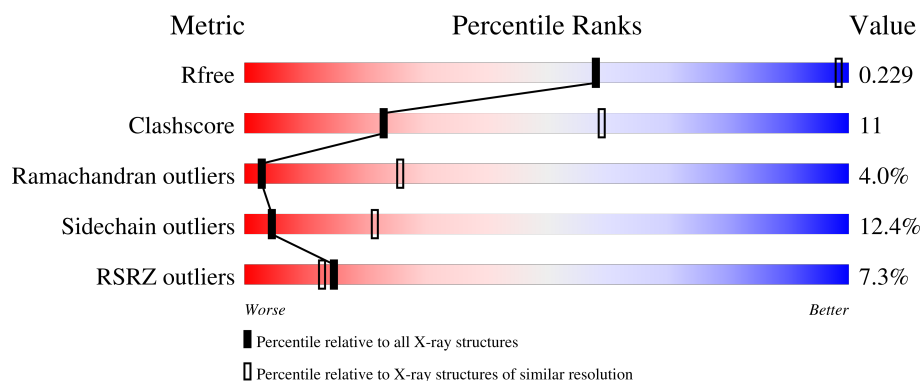
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 4.60 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	1062 (5.40-3.80)
Clashscore	141614	1130 (5.40-3.80)
Ramachandran outliers	138981	1074 (5.40-3.80)
Sidechain outliers	138945	1055 (5.40-3.80)
RSRZ outliers	127900	1113 (5.50-3.70)

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	1382	<div> <div>5%</div> <div> <div></div> <div>58%</div> <div>28%</div> <div>•</div> <div>9%</div> </div> </div>
1	B	1382	<div> <div>9%</div> <div> <div></div> <div>58%</div> <div>28%</div> <div>•</div> <div>9%</div> </div> </div>

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 20112 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 UDP-glucose-glycoprotein glucosyltransferase-like protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1254	Total	C	N	O	S	0	0	0
			10056	6439	1711	1879	27			
1	B	1254	Total	C	N	O	S	0	0	0
			10056	6439	1711	1879	27			

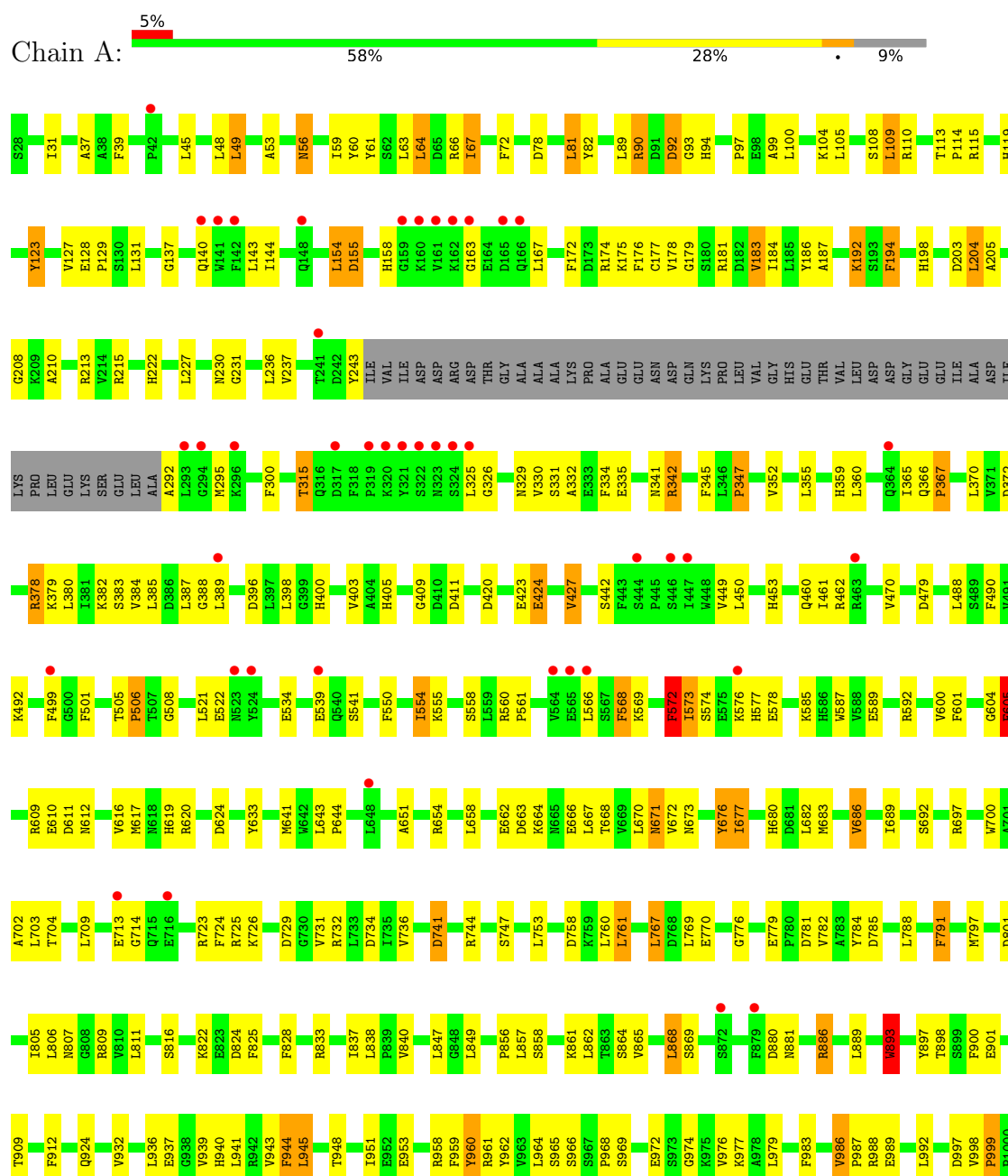
There are 4 discrepancies between the modelled and reference sequences:

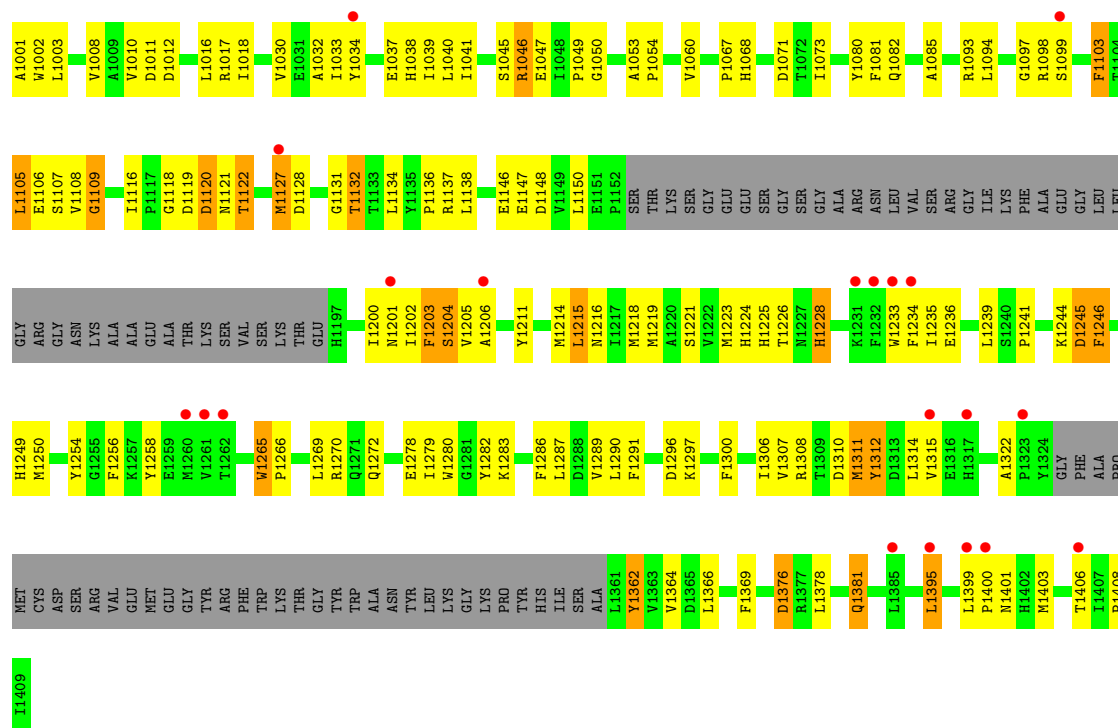
Chain	Residue	Modelled	Actual	Comment	Reference
A	177	CYS	GLY	engineered mutation	UNP G0SB58
A	786	CYS	ALA	engineered mutation	UNP G0SB58
B	177	CYS	GLY	engineered mutation	UNP G0SB58
B	786	CYS	ALA	engineered mutation	UNP G0SB58

3 Residue-property plots [i](#)

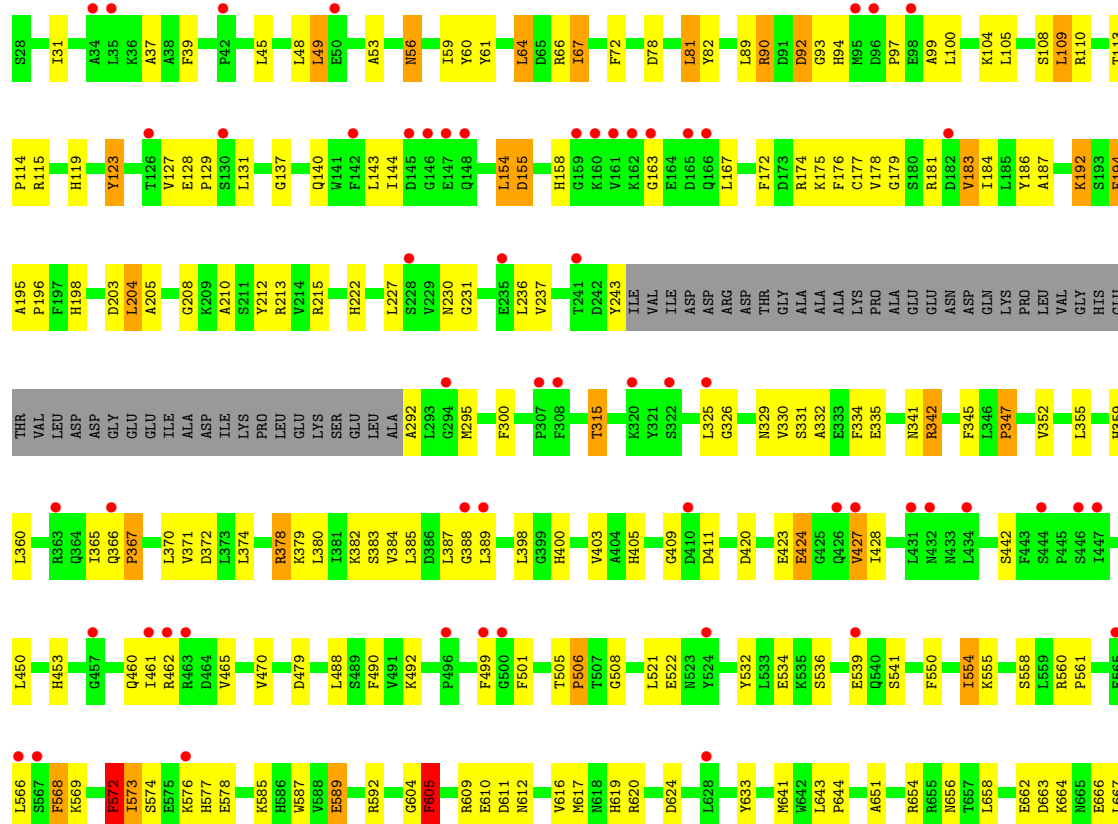
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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: UDP-glucose-glycoprotein glucosyltransferase-like protein





• Molecule 1: UDP-glucose-glycoprotein glucosyltransferase-like protein



THR	GLY	TYR	ALA	ASN	TYR	LEU	LYS	LYS	PRO	TYR	HIS	ILE	SER	ALA	L1361	L1362	Y1362	Y1363	V1364	D1365	L1366	Y1369	F1369	D1376	R1377	L1378	Q1381	S1391	L1392	A1393	N1394	L1395	D1398	L1399	P1400	N1401	H1402	M1403	T1406	I1407	P1408	I1409																
W1265	P1266	L1269	R1270	Q1271	Q1272	E1278	S1204	W1279	W1280	G1281	Y1282	K1283	F1286	L1287	D1288	V1289	L1290	F1291	D1296	K1297	F1300	I1306	V1307	R1308	T1309	D1310	M1311	Y1312	D1313	L1314	V1315	A1322	P1323	Y1324	GLY	PHE	ALA	PRO	MET	CYS	ASP	SER	ARG	VAL	GLU	GLY	TYR	GLY	ARG	PHE	TRP	LYS						
THR	GLU	H1197	I1200	N1201	I1202	F1203	S1204	V1205	A1206	Y1211	M1214	L1215	M1216	I1217	M1218	M1219	A1220	S1221	V1222	M1223	H1224	H1225	T1226	M1227	H1228	T1229	V1230	K1231	F1232	W1233	F1234	I1235	E1236	Q1237	F1238	L1239	S1240	P1241	K1244	D1245	F1246	H1249	M1250	Y1254	G1255	F1256	K1257	Y1258	E1259	M1261	V1262	L1263	L1264	L1265				
F1129	Q1130	T1132	T1133	L1134	Y1135	P1136	R1137	L1138	E1146	E1147	D1148	V1149	L1150	E1151	P1152	SER	THR	LYS	SER	GLY	GLU	GLU	SER	GLY	GLY	ALA	ARG	ASN	LEU	VAL	SER	ARG	GLY	ILE	LYS	PHE	ALA	GLU	GLY	LEU	LEU	GLY	ARG	GLY	ASN	LYS	ALA	ALA	GLU	GLU	THR	LYS	SER	VAL	LYS	SER	THR	LYS
I1041	S1045	R1046	E1047	I1048	P1049	G1050	P1054	G1056	V1057	Q1058	V1060	P1067	H1068	D1071	T1072	I1073	Y1080	F1081	Q1082	F1083	R1084	A1085	V1088	R1093	L1094	G1097	R1098	S1099	I1102	F1103	T1104	L1105	E1106	S1107	V1108	G1109	P1115	I1116	G1118	D1119	D1120	N1121	T1122	M1127	D1128	E1037	H1038	T1039	L1040									
R958	F959	Y960	R961	Y962	Y963	L964	S965	S966	S967	P968	S969	E972	S973	G974	K975	Y976	K977	A978	L979	S980	F983	Y986	P987	F988	E989	L992	Y996	D997	Y998	P999	F1000	A1001	W1002	L1003	V1008	D1011	D1012	L1016	R1017	I1018	V1030	E1031	A1032	I1033	Y1034	E1037	H1038	T1039	L1040									
T668	F669	L670	N671	V672	N673	Y676	L677	H680	L682	M683	V686	L689	S692	R697	W700	A701	A702	L703	T704	L709	E713	G714	Q715	R723	F724	R725	K726	D729	G730	V731	R732	L733	D734	I735	V736	D741	R744	S747	L753	D758	K759	L760																
L761	D767	L768	E770	G776	E777	L778	E779	D781	V782	A783	D785	L788	F791	M797	D801	I805	L806	F900	N807	G808	R809	L811	S816	K822	E823	D824	F825	L936	E937	G938	V939	H940	L941	R942	L838	P839	F944	L945	T948	L847	G848	L849	P856	L857	S858	R861	L862											

4 Data and refinement statistics

Property	Value	Source
Space group	P 43	Depositor
Cell constants a, b, c, α , β , γ	139.05Å 139.05Å 176.09Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	139.04 – 4.60 139.05 – 4.77	Depositor EDS
% Data completeness (in resolution range)	32.1 (139.04-4.60) 36.0 (139.05-4.77)	Depositor EDS
R_{merge}	0.28	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.76 (at 4.88Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.203 , 0.234 0.202 , 0.229	Depositor DCC
R_{free} test set	339 reflections (5.63%)	wwPDB-VP
Wilson B-factor (Å ²)	206.4	Xtriage
Anisotropy	0.047	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.39 , 500.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.39$, $\langle L^2 \rangle = 0.22$	Xtriage
Estimated twinning fraction	0.369 for h,-k,-l	Xtriage
F_o, F_c correlation	0.90	EDS
Total number of atoms	20112	wwPDB-VP
Average B, all atoms (Å ²)	173.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.41% 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

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	0.64	0/10293	0.77	0/13959
1	B	0.63	0/10293	0.76	0/13959
All	All	0.63	0/20586	0.76	0/27918

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

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	10056	0	9955	226	0
1	B	10056	0	9955	231	0
All	All	20112	0	19910	457	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 457 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:1109:GLY:H	1:A:1136:PRO:HA	1.12	1.11
1:B:1109:GLY:H	1:B:1136:PRO:HA	1.12	1.06

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:382:LYS:HA	1:A:385:LEU:HG	1.51	0.93
1:B:382:LYS:HA	1:B:385:LEU:HG	1.51	0.92
1:B:833:ARG:HA	1:B:837:ILE:HB	1.55	0.87

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	1246/1382 (90%)	988 (79%)	208 (17%)	50 (4%)	3	26
1	B	1246/1382 (90%)	989 (79%)	207 (17%)	50 (4%)	3	26
All	All	2492/2764 (90%)	1977 (79%)	415 (17%)	100 (4%)	3	26

5 of 100 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	155	ASP
1	A	424	GLU
1	A	427	VAL
1	A	506	PRO
1	A	541	SER

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	1097/1197 (92%)	961 (88%)	136 (12%)	4	21
1	B	1097/1197 (92%)	960 (88%)	137 (12%)	4	21
All	All	2194/2394 (92%)	1921 (88%)	273 (12%)	4	21

5 of 273 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	1012	ASP
1	B	1071	ASP
1	B	1254	TYR
1	A	1037	GLU
1	A	1012	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 40 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	656	ASN
1	B	1082	GLN
1	B	728	ASN
1	B	946	ASN
1	B	1227	ASN

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 monosaccharides in this entry.

5.6 Ligand geometry ⓘ

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 [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	1254/1382 (90%)	0.24	63 (5%) 28 25	73, 141, 251, 333	0
1	B	1254/1382 (90%)	0.50	121 (9%) 8 8	106, 188, 295, 398	0
All	All	2508/2764 (90%)	0.37	184 (7%) 15 13	73, 168, 272, 398	0

The worst 5 of 184 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	160	LYS	9.3
1	A	565	GLU	7.7
1	B	565	GLU	7.6
1	B	159	GLY	7.2
1	A	160	LYS	6.9

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](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.