



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 9, 2017 – 01:08 PM EDT

PDB ID : 2IQI
Title : Crystal structure of protein XCC0632 from *Xanthomonas campestris*, Pfam DUF330
Authors : Bonanno, J.B.; Gilmore, J.; Bain, K.T.; McKenzie, C.; Pelletier, L.; Wasserman, S.; Burley, S.K.; Almo, S.C.; New York SGX Research Center for Structural Genomics (NYSGXRC)
Deposited on : unknown
Resolution : 2.70 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

<http://wwpdb.org/validation/2016/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)	:	1.9-1692
EDS	:	rb-20030345
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20030345

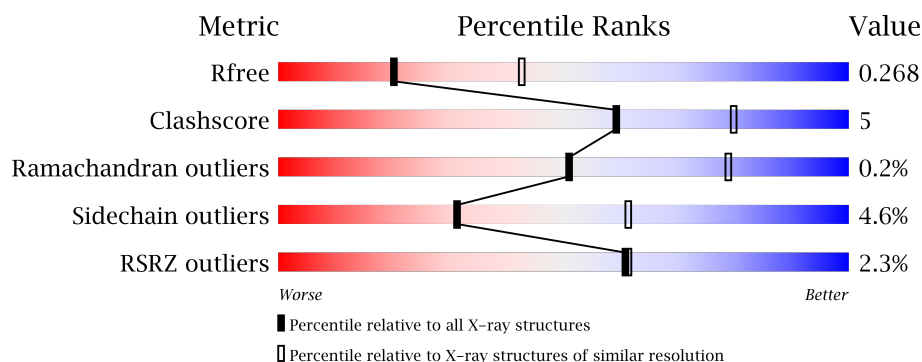
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.70 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}	100719	2259 (2.70-2.70)
Clashscore	112137	2590 (2.70-2.70)
Ramachandran outliers	110173	2550 (2.70-2.70)
Sidechain outliers	110143	2550 (2.70-2.70)
RSRZ outliers	101464	2275 (2.70-2.70)

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 on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. 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	192	<div> <div>3%</div> <div> <div></div> <div>78%</div> <div>11%</div> <div>10%</div> </div> </div>
1	B	192	<div> <div>3%</div> <div> <div></div> <div>78%</div> <div>9%</div> <div>11%</div> </div> </div>
1	C	192	<div> <div>3%</div> <div> <div></div> <div>73%</div> <div>16%</div> <div>11%</div> </div> </div>
1	D	192	<div> <div>0%</div> <div> <div></div> <div>80%</div> <div>9%</div> <div>11%</div> </div> </div>
1	E	192	<div> <div>2%</div> <div> <div></div> <div>76%</div> <div>11%</div> <div>11%</div> </div> </div>

Continued on next page...

Mol	Chain	Length	Quality of chain
1	F	192	<p>73% 17% 9%</p>
1	G	192	<p>76% 14% 10%</p>
1	H	192	<p>77% 11% 11%</p>

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 10680 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 Hypothetical protein XCC0632.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	173	Total	C	N	O	S	0	0	0
			1316	827	232	256	1			
1	B	171	Total	C	N	O	S	0	0	0
			1303	818	230	254	1			
1	C	171	Total	C	N	O	S	0	0	0
			1303	818	230	254	1			
1	D	171	Total	C	N	O	S	0	0	0
			1303	818	230	254	1			
1	E	171	Total	C	N	O	S	0	0	0
			1303	818	230	254	1			
1	F	175	Total	C	N	O	S	0	0	0
			1335	839	237	258	1			
1	G	172	Total	C	N	O	S	0	0	0
			1311	824	231	255	1			
1	H	171	Total	C	N	O	S	0	0	0
			1303	818	230	254	1			

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	23	SER	-	CLONING ARTIFACT	UNP Q8PCT0
A	24	LEU	-	CLONING ARTIFACT	UNP Q8PCT0
A	25	SER	-	CLONING ARTIFACT	UNP Q8PCT0
A	26	LEU	-	CLONING ARTIFACT	UNP Q8PCT0
B	23	SER	-	CLONING ARTIFACT	UNP Q8PCT0
B	24	LEU	-	CLONING ARTIFACT	UNP Q8PCT0
B	25	SER	-	CLONING ARTIFACT	UNP Q8PCT0
B	26	LEU	-	CLONING ARTIFACT	UNP Q8PCT0
C	23	SER	-	CLONING ARTIFACT	UNP Q8PCT0
C	24	LEU	-	CLONING ARTIFACT	UNP Q8PCT0
C	25	SER	-	CLONING ARTIFACT	UNP Q8PCT0
C	26	LEU	-	CLONING ARTIFACT	UNP Q8PCT0
D	23	SER	-	CLONING ARTIFACT	UNP Q8PCT0

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	24	LEU	-	CLONING ARTIFACT	UNP Q8PCT0
D	25	SER	-	CLONING ARTIFACT	UNP Q8PCT0
D	26	LEU	-	CLONING ARTIFACT	UNP Q8PCT0
E	23	SER	-	CLONING ARTIFACT	UNP Q8PCT0
E	24	LEU	-	CLONING ARTIFACT	UNP Q8PCT0
E	25	SER	-	CLONING ARTIFACT	UNP Q8PCT0
E	26	LEU	-	CLONING ARTIFACT	UNP Q8PCT0
F	23	SER	-	CLONING ARTIFACT	UNP Q8PCT0
F	24	LEU	-	CLONING ARTIFACT	UNP Q8PCT0
F	25	SER	-	CLONING ARTIFACT	UNP Q8PCT0
F	26	LEU	-	CLONING ARTIFACT	UNP Q8PCT0
G	23	SER	-	CLONING ARTIFACT	UNP Q8PCT0
G	24	LEU	-	CLONING ARTIFACT	UNP Q8PCT0
G	25	SER	-	CLONING ARTIFACT	UNP Q8PCT0
G	26	LEU	-	CLONING ARTIFACT	UNP Q8PCT0
H	23	SER	-	CLONING ARTIFACT	UNP Q8PCT0
H	24	LEU	-	CLONING ARTIFACT	UNP Q8PCT0
H	25	SER	-	CLONING ARTIFACT	UNP Q8PCT0
H	26	LEU	-	CLONING ARTIFACT	UNP Q8PCT0

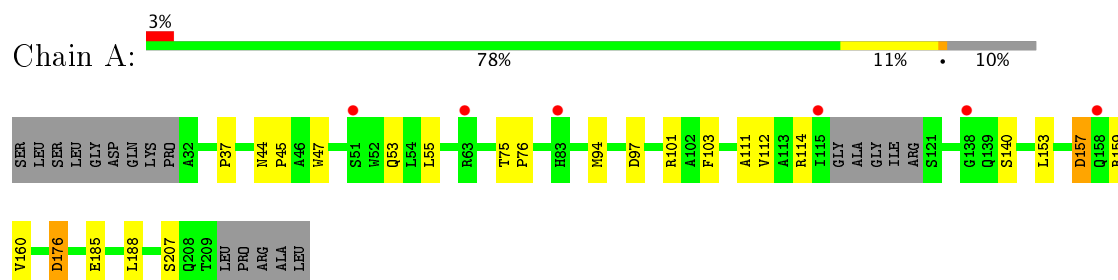
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	30	Total O 30 30	0	0
2	B	20	Total O 20 20	0	0
2	C	21	Total O 21 21	0	0
2	D	28	Total O 28 28	0	0
2	E	23	Total O 23 23	0	0
2	F	21	Total O 21 21	0	0
2	G	29	Total O 29 29	0	0
2	H	31	Total O 31 31	0	0

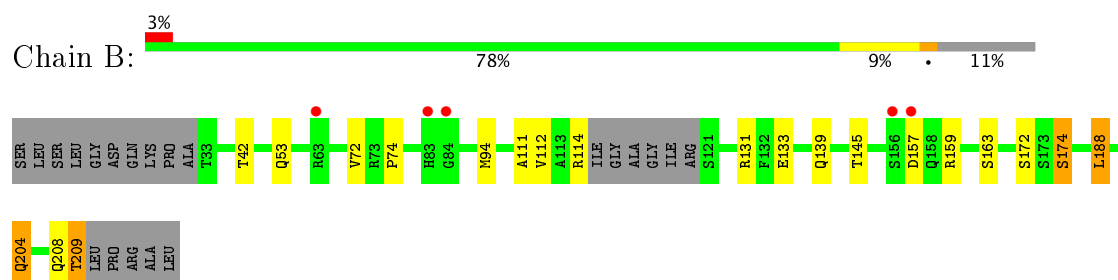
3 Residue-property plots [i](#)

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.

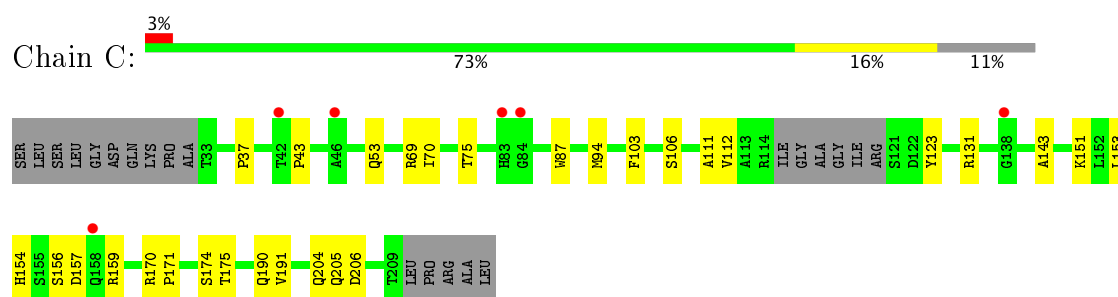
- Molecule 1: Hypothetical protein XCC0632



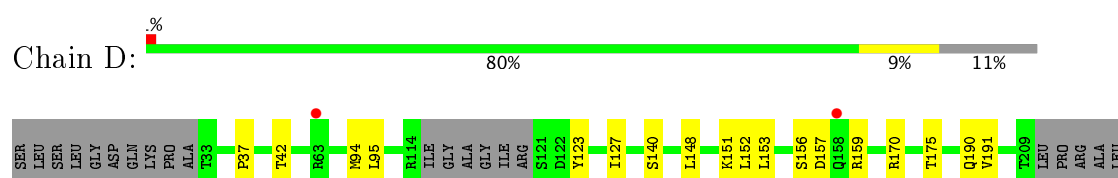
- Molecule 1: Hypothetical protein XCC0632



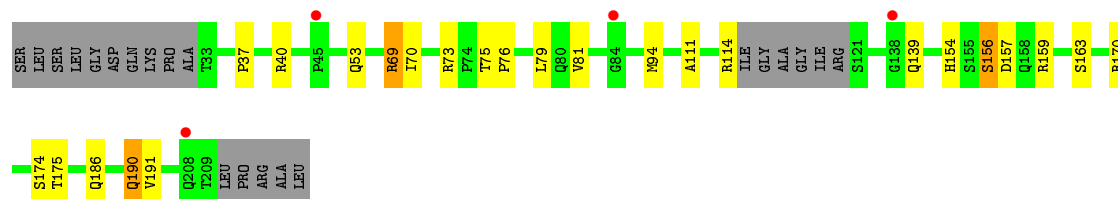
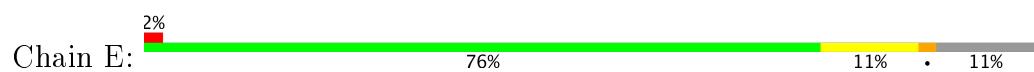
- Molecule 1: Hypothetical protein XCC0632



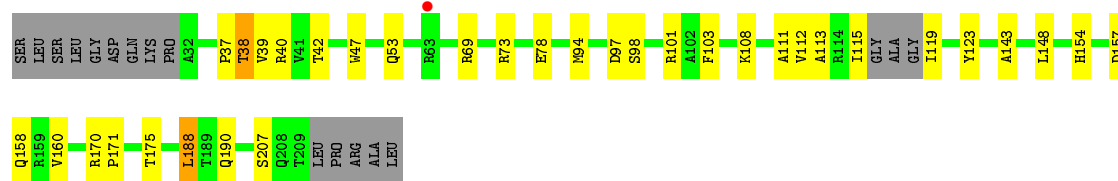
- Molecule 1: Hypothetical protein XCC0632



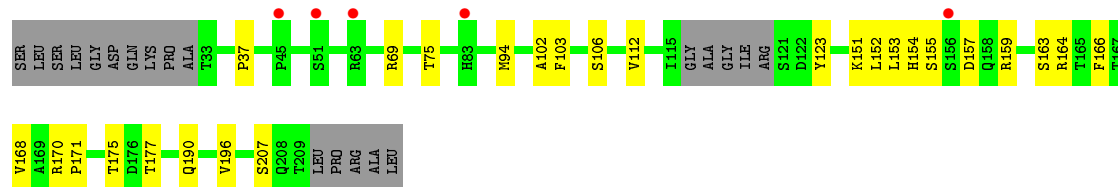
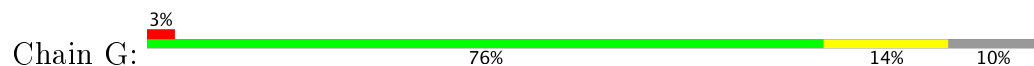
- Molecule 1: Hypothetical protein XCC0632



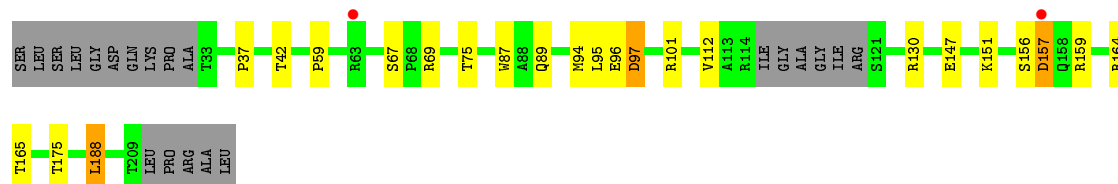
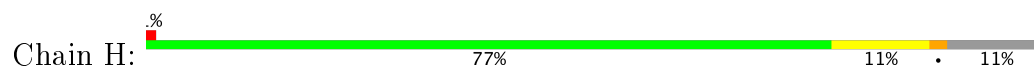
• Molecule 1: Hypothetical protein XCC0632



• Molecule 1: Hypothetical protein XCC0632



• Molecule 1: Hypothetical protein XCC0632



4 Data and refinement statistics

Property	Value	Source
Space group	P 2 ₁ 2 ₁ 2	Depositor
Cell constants a, b, c, α , β , γ	155.93Å 103.11Å 121.65Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.89 – 2.70 19.89 – 2.70	Depositor EDS
% Data completeness (in resolution range)	100.0 (19.89-2.70) 100.0 (19.89-2.70)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	0.15	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.93 (at 2.71Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.209 , 0.275 0.208 , 0.268	Depositor DCC
R_{free} test set	1548 reflections (2.93%)	DCC
Wilson B-factor (Å ²)	39.6	Xtriage
Anisotropy	0.144	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.36 , 36.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	10680	wwPDB-VP
Average B, all atoms (Å ²)	46.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.69% 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](#)

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	0.79	0/1343	0.83	1/1839 (0.1%)
1	B	0.71	0/1330	0.77	0/1821
1	C	0.73	0/1330	0.80	0/1821
1	D	0.78	0/1330	0.82	1/1821 (0.1%)
1	E	0.76	1/1330 (0.1%)	0.82	3/1821 (0.2%)
1	F	0.83	1/1362 (0.1%)	0.83	1/1864 (0.1%)
1	G	0.75	0/1338	0.76	0/1832
1	H	0.77	0/1330	0.83	0/1821
All	All	0.77	2/10693 (0.0%)	0.81	6/14640 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	190	GLN	CG-CD	5.70	1.64	1.51
1	F	78	GLU	CD-OE1	5.04	1.31	1.25

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	176	ASP	CB-CG-OD1	5.87	123.58	118.30
1	F	73	ARG	NE-CZ-NH2	-5.54	117.53	120.30
1	E	69	ARG	CG-CD-NE	5.43	123.21	111.80
1	E	73	ARG	NE-CZ-NH2	-5.24	117.68	120.30
1	D	95	LEU	CB-CG-CD2	-5.14	102.27	111.00

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	1316	0	1306	15	0
1	B	1303	0	1290	11	0
1	C	1303	0	1290	20	0
1	D	1303	0	1290	10	0
1	E	1303	0	1290	13	0
1	F	1335	0	1330	20	0
1	G	1311	0	1301	15	0
1	H	1303	0	1290	13	0
2	A	30	0	0	0	0
2	B	20	0	0	1	0
2	C	21	0	0	0	0
2	D	28	0	0	1	0
2	E	23	0	0	1	0
2	F	21	0	0	0	0
2	G	29	0	0	1	0
2	H	31	0	0	1	0
All	All	10680	0	10387	112	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 5.

The worst 5 of 112 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:C:157:ASP:OD2	1:C:159:ARG:HB2	1.66	0.94
1:H:164:ARG:HD3	2:H:216:HOH:O	1.69	0.91
1:B:94:MET:HE2	1:B:188:LEU:HD11	1.54	0.87
1:G:123:TYR:HB3	1:G:152:LEU:HD11	1.58	0.83
1:E:170:ARG:NH1	1:E:190:GLN:OE1	2.14	0.81

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

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	169/192 (88%)	162 (96%)	7 (4%)	0	100	100
1	B	167/192 (87%)	162 (97%)	4 (2%)	1 (1%)	28	56
1	C	167/192 (87%)	161 (96%)	5 (3%)	1 (1%)	28	56
1	D	167/192 (87%)	160 (96%)	7 (4%)	0	100	100
1	E	167/192 (87%)	160 (96%)	6 (4%)	1 (1%)	28	56
1	F	171/192 (89%)	163 (95%)	8 (5%)	0	100	100
1	G	168/192 (88%)	163 (97%)	5 (3%)	0	100	100
1	H	167/192 (87%)	160 (96%)	7 (4%)	0	100	100
All	All	1343/1536 (87%)	1291 (96%)	49 (4%)	3 (0%)	51	79

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	139	GLN
1	E	139	GLN
1	C	205	GLN

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	140/154 (91%)	135 (96%)	5 (4%)	40	70
1	B	139/154 (90%)	132 (95%)	7 (5%)	28	57
1	C	139/154 (90%)	134 (96%)	5 (4%)	40	70
1	D	139/154 (90%)	135 (97%)	4 (3%)	48	77
1	E	139/154 (90%)	132 (95%)	7 (5%)	28	57
1	F	142/154 (92%)	134 (94%)	8 (6%)	25	51
1	G	140/154 (91%)	134 (96%)	6 (4%)	33	64
1	H	139/154 (90%)	130 (94%)	9 (6%)	20	44

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	1117/1232 (91%)	1066 (95%)	51 (5%)	31 61

5 of 51 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	156	SER
1	F	38	THR
1	H	156	SER
1	E	174	SER
1	F	42	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	53	GLN
1	E	154	HIS
1	F	186	GLN
1	D	186	GLN
1	F	154	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 ⓘ

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, 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	173/192 (90%)	-0.16	6 (3%)	44	44	31, 43, 57, 65	0
1	B	171/192 (89%)	0.05	5 (2%)	52	52	37, 49, 67, 75	0
1	C	171/192 (89%)	0.02	6 (3%)	44	44	36, 47, 66, 83	0
1	D	171/192 (89%)	-0.12	2 (1%)	79	80	33, 45, 57, 67	0
1	E	171/192 (89%)	0.03	4 (2%)	61	61	36, 46, 62, 72	0
1	F	175/192 (91%)	-0.14	1 (0%)	89	90	31, 41, 57, 67	0
1	G	172/192 (89%)	-0.01	5 (2%)	52	52	36, 46, 61, 72	0
1	H	171/192 (89%)	-0.14	2 (1%)	79	80	32, 44, 58, 68	0
All	All	1375/1536 (89%)	-0.06	31 (2%)	61	61	31, 45, 62, 83	0

The worst 5 of 31 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	138	GLY	3.4
1	B	63	ARG	3.4
1	F	63	ARG	3.2
1	E	84	GLY	3.1
1	G	51	SER	3.0

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.