



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 31, 2025 – 09:58 PM JST

PDB ID : 8YZQ / pdb_00008yzq
Title : Toxascaris leonina galectin (Tl-gal) W77F/W212F oligomer
Authors : Ha, M.S.; Jang, S.B.; Jeong, M.S.
Deposited on : 2024-04-08
Resolution : 2.15 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)	:	2.0rc1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.006 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.42

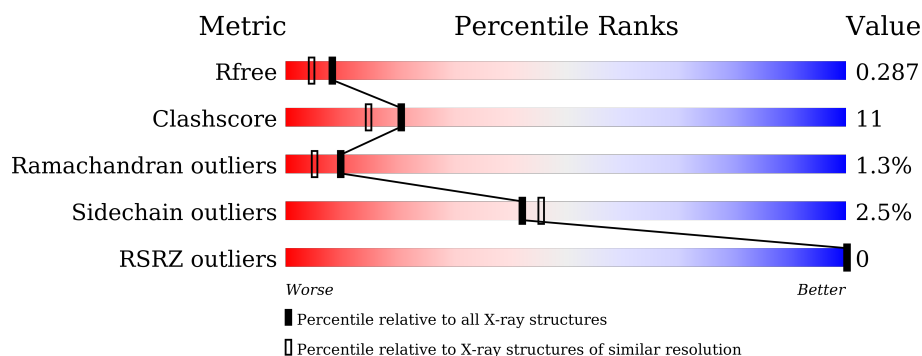
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.15 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}	164625	1881 (2.16-2.16)
Clashscore	180529	2047 (2.16-2.16)
Ramachandran outliers	177936	2027 (2.16-2.16)
Sidechain outliers	177891	2026 (2.16-2.16)
RSRZ outliers	164620	1882 (2.16-2.16)

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	278	
1	B	278	
1	C	278	
1	D	278	
1	E	278	
1	F	278	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	278	 73% 22% . .
1	H	278	 76% 18% . .

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 17893 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 Galectin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	266	Total	C	N	O	S	0	0	0
			2136	1361	369	404	2			
1	B	267	Total	C	N	O	S	0	0	0
			2144	1365	371	406	2			
1	C	269	Total	C	N	O	S	0	0	0
			2160	1374	373	411	2			
1	D	267	Total	C	N	O	S	0	0	0
			2144	1365	371	406	2			
1	E	269	Total	C	N	O	S	0	0	0
			2157	1372	373	410	2			
1	F	267	Total	C	N	O	S	0	0	0
			2144	1365	371	406	2			
1	G	268	Total	C	N	O	S	0	0	0
			2151	1369	372	408	2			
1	H	266	Total	C	N	O	S	0	0	0
			2136	1361	369	404	2			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	77	PHE	TRP	engineered mutation	UNP A0A1L1QJZ7
A	212	PHE	TRP	engineered mutation	UNP A0A1L1QJZ7
B	77	PHE	TRP	engineered mutation	UNP A0A1L1QJZ7
B	212	PHE	TRP	engineered mutation	UNP A0A1L1QJZ7
C	77	PHE	TRP	engineered mutation	UNP A0A1L1QJZ7
C	212	PHE	TRP	engineered mutation	UNP A0A1L1QJZ7
D	77	PHE	TRP	engineered mutation	UNP A0A1L1QJZ7
D	212	PHE	TRP	engineered mutation	UNP A0A1L1QJZ7
E	77	PHE	TRP	engineered mutation	UNP A0A1L1QJZ7
E	212	PHE	TRP	engineered mutation	UNP A0A1L1QJZ7
F	77	PHE	TRP	engineered mutation	UNP A0A1L1QJZ7
F	212	PHE	TRP	engineered mutation	UNP A0A1L1QJZ7
G	77	PHE	TRP	engineered mutation	UNP A0A1L1QJZ7

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
G	212	PHE	TRP	engineered mutation	UNP A0A1L1QJZ7
H	77	PHE	TRP	engineered mutation	UNP A0A1L1QJZ7
H	212	PHE	TRP	engineered mutation	UNP A0A1L1QJZ7

- Molecule 2 is water.

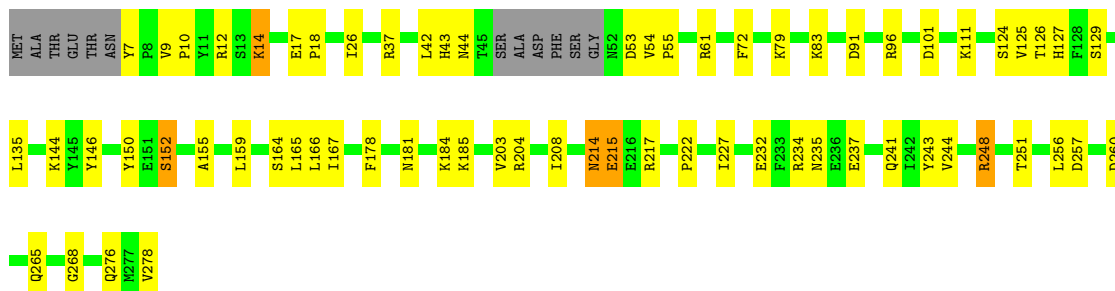
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	97	Total O 97 97	0	0
2	B	109	Total O 109 109	0	0
2	C	88	Total O 88 88	0	0
2	D	104	Total O 104 104	0	0
2	E	81	Total O 81 81	0	0
2	F	81	Total O 81 81	0	0
2	G	79	Total O 79 79	0	0
2	H	82	Total O 82 82	0	0

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.

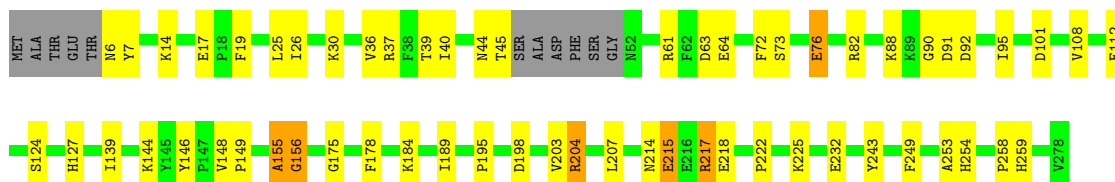
- Molecule 1: Galectin

Chain A: 



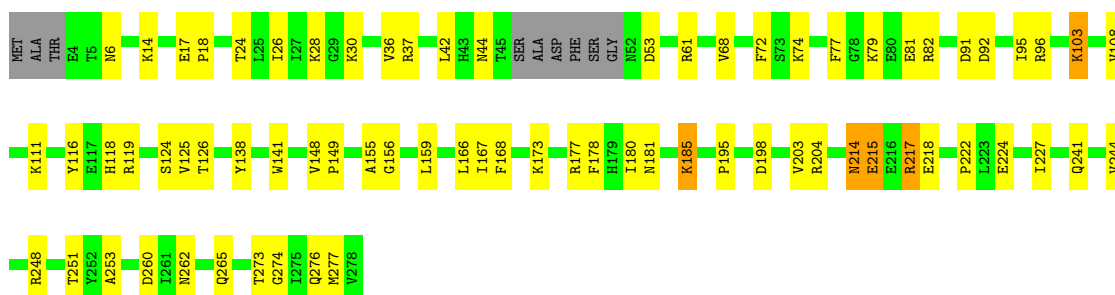
- Molecule 1: Galectin

Chain B: 




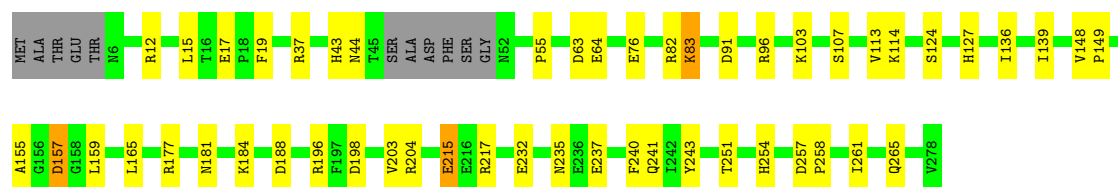
- Molecule 1: Galectin

Chain C: 



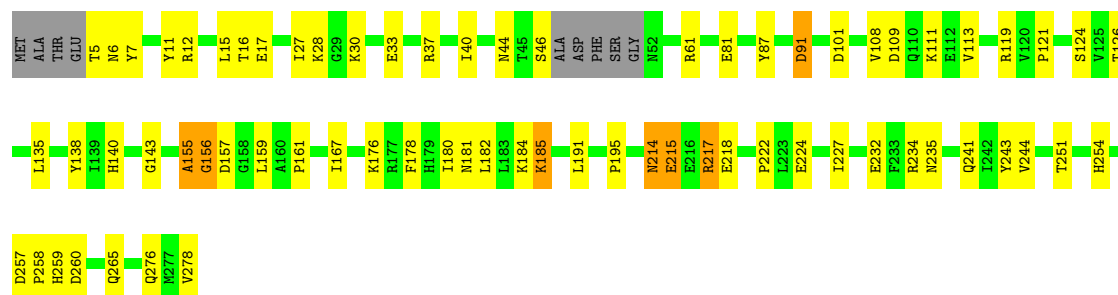
- Molecule 1: Galectin

Chain D:  78% 17%



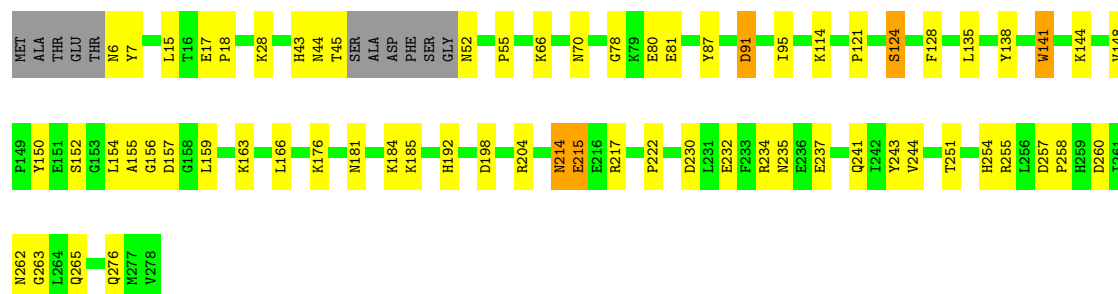
• Molecule 1: Galectin

Chain E:  72% 23%



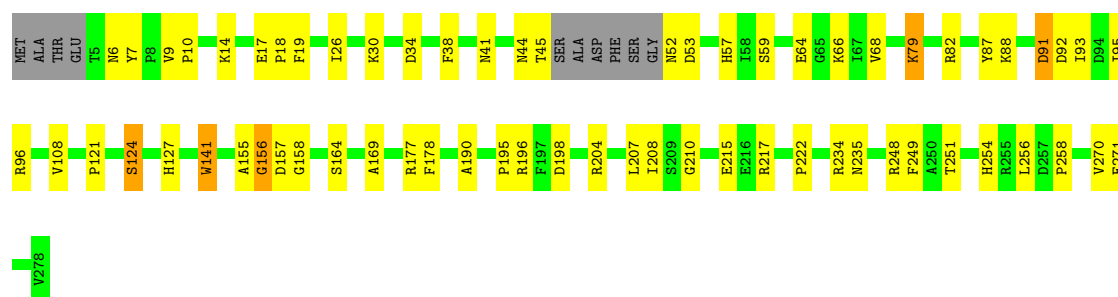
• Molecule 1: Galectin

Chain F:  72% 22%



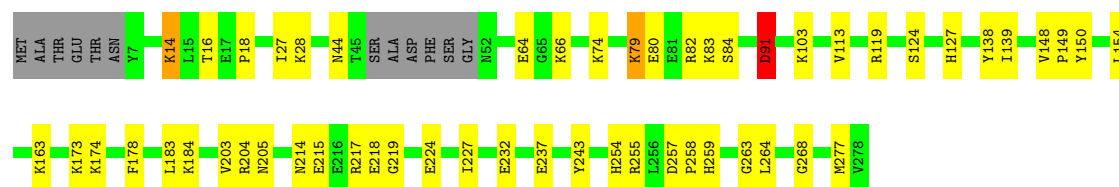
• Molecule 1: Galectin

Chain G:  73% 22%



• Molecule 1: Galectin

76% 18% . .



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	78.52Å 82.15Å 114.59Å 90.02° 90.03° 90.01°	Depositor
Resolution (Å)	28.65 – 2.15 28.65 – 2.15	Depositor EDS
% Data completeness (in resolution range)	89.4 (28.65-2.15) 89.4 (28.65-2.15)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	7.21 (at 2.16Å)	Xtriage
Refinement program	PHENIX 1.14 _3260	Depositor
R, R_{free}	0.234 , 0.286 0.235 , 0.287	Depositor DCC
R_{free} test set	6842 reflections (4.92%)	wwPDB-VP
Wilson B-factor (Å ²)	34.4	Xtriage
Anisotropy	0.138	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.37 , 19.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.44$, $\langle L^2 \rangle = 0.26$	Xtriage
Estimated twinning fraction	0.047 for k,-h,l 0.047 for -k,h,l 0.407 for h,-k,-l 0.409 for -h,k,-l 0.447 for -h,-k,l 0.048 for k,h,-l 0.047 for -k,-h,-l	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	17893	wwPDB-VP
Average B, all atoms (Å ²)	36.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.23% 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	0.50	1/2187 (0.0%)	0.63	0/2947
1	B	0.50	1/2195 (0.0%)	0.63	0/2958
1	C	0.45	0/2211	0.66	0/2980
1	D	0.47	0/2195	0.64	1/2958 (0.0%)
1	E	0.42	0/2208	0.62	0/2976
1	F	0.48	1/2195 (0.0%)	0.63	1/2958 (0.0%)
1	G	0.44	0/2202	0.64	3/2968 (0.1%)
1	H	0.43	0/2187	0.62	1/2947 (0.0%)
All	All	0.46	3/17580 (0.0%)	0.63	6/23692 (0.0%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	14	LYS	CE-NZ	-6.90	1.31	1.49
1	F	141	TRP	CG-CD1	6.15	1.45	1.36
1	B	76	GLU	CD-OE2	5.61	1.31	1.25

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	G	79	LYS	CD-CE-NZ	-6.35	97.08	111.70
1	G	79	LYS	CA-CB-CG	-6.30	99.53	113.40
1	F	141	TRP	CH2-CZ2-CE2	-5.96	111.44	117.40
1	G	91	ASP	CB-CG-OD1	5.73	123.46	118.30
1	H	91	ASP	CB-CG-OD1	5.31	123.08	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	2136	0	2090	53	1
1	B	2144	0	2096	47	1
1	C	2160	0	2109	70	0
1	D	2144	0	2095	32	0
1	E	2157	0	2108	47	0
1	F	2144	0	2096	50	0
1	G	2151	0	2103	44	0
1	H	2136	0	2089	36	0
2	A	97	0	0	11	2
2	B	109	0	0	15	1
2	C	88	0	0	16	0
2	D	104	0	0	5	1
2	E	81	0	0	11	0
2	F	81	0	0	13	4
2	G	79	0	0	9	2
2	H	82	0	0	8	2
All	All	17893	0	16786	374	7

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 374 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:37:ARG:HH22	1:C:61:ARG:CZ	1.16	1.58
1:C:37:ARG:NH2	1:C:61:ARG:NE	1.66	1.37
1:C:37:ARG:NH2	1:C:61:ARG:CZ	1.88	1.37
1:A:14:LYS:HE3	1:A:127:HIS:CD2	1.76	1.19
1:G:57:HIS:NE2	2:G:302:HOH:O	1.89	1.05

The worst 5 of 7 symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:F:358:HOH:O	2:H:373:HOH:O[1_655]	2.08	0.12
2:F:355:HOH:O	2:H:373:HOH:O[1_655]	2.10	0.10
2:A:362:HOH:O	2:D:386:HOH:O[1_454]	2.11	0.09
1:A:14:LYS:CE	1:B:76:GLU:OE2[1_444]	2.13	0.07
2:F:317:HOH:O	2:G:319:HOH:O[1_565]	2.14	0.06

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	262/278 (94%)	251 (96%)	8 (3%)	3 (1%)	12	7
1	B	263/278 (95%)	257 (98%)	3 (1%)	3 (1%)	12	7
1	C	265/278 (95%)	253 (96%)	8 (3%)	4 (2%)	8	4
1	D	263/278 (95%)	255 (97%)	5 (2%)	3 (1%)	12	7
1	E	265/278 (95%)	252 (95%)	8 (3%)	5 (2%)	6	2
1	F	263/278 (95%)	254 (97%)	4 (2%)	5 (2%)	6	2
1	G	264/278 (95%)	254 (96%)	6 (2%)	4 (2%)	8	4
1	H	262/278 (94%)	253 (97%)	8 (3%)	1 (0%)	30	27
All	All	2107/2224 (95%)	2029 (96%)	50 (2%)	28 (1%)	10	5

5 of 28 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	155	ALA
1	A	214	ASN
1	A	215	GLU
1	B	155	ALA
1	C	215	GLU

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	231/240 (96%)	226 (98%)	5 (2%)	47	51
1	B	232/240 (97%)	228 (98%)	4 (2%)	56	61
1	C	234/240 (98%)	227 (97%)	7 (3%)	36	37
1	D	232/240 (97%)	228 (98%)	4 (2%)	56	61
1	E	234/240 (98%)	226 (97%)	8 (3%)	32	31
1	F	232/240 (97%)	227 (98%)	5 (2%)	47	51
1	G	233/240 (97%)	226 (97%)	7 (3%)	36	37
1	H	231/240 (96%)	224 (97%)	7 (3%)	36	37
All	All	1859/1920 (97%)	1812 (98%)	47 (2%)	42	45

5 of 47 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	66	LYS
1	G	124	SER
1	F	124	SER
1	F	234	ARG
1	G	157	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	127	HIS
1	E	140	HIS

5.3.3 RNA ⓘ

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 oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

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	266/278 (95%)	-1.16	0 100 100	24, 34, 47, 66	0
1	B	267/278 (96%)	-1.19	0 100 100	23, 33, 48, 63	0
1	C	269/278 (96%)	-1.17	0 100 100	24, 34, 47, 81	0
1	D	267/278 (96%)	-1.20	0 100 100	23, 32, 47, 70	0
1	E	269/278 (96%)	-1.04	0 100 100	26, 38, 52, 79	0
1	F	267/278 (96%)	-1.03	0 100 100	28, 38, 52, 67	0
1	G	268/278 (96%)	-0.99	0 100 100	27, 37, 51, 73	0
1	H	266/278 (95%)	-1.05	0 100 100	25, 36, 50, 64	0
All	All	2139/2224 (96%)	-1.10	0 100 100	23, 35, 50, 81	0

There are no RSRZ outliers to report.

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.