WORLDWIDE wwPDB PROTEIN DATA BANK

Full wwPDB X-ray Structure Validation Report (i)

May 22, 2024 – 06:12 PM EDT

PDB ID	:	8W04
Title	:	Crystal structure of DUF1735-domain containing protein (GH18-like) from Bacteroides faecium at 2.9 A resolution (Space group P21)
Authors	:	Sastre, D.E.; Navarro, M.V.A.S.; Sultana, N.; Sundberg, E.J.
Deposited on	:	2024-02-13
Resolution	:	2.91 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

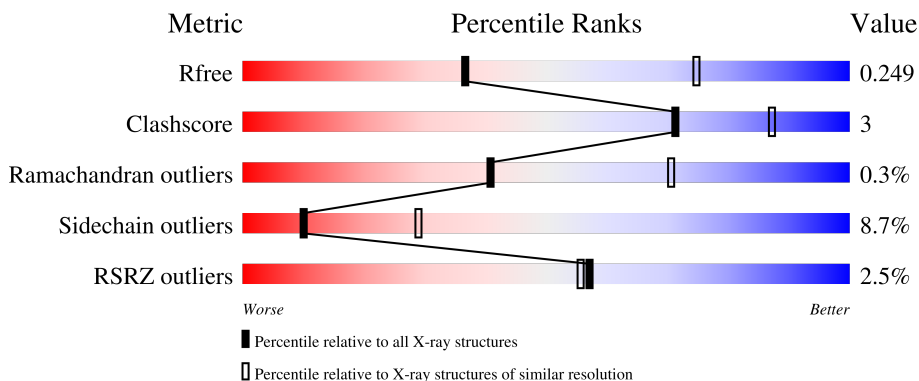|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 2.91 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	Percentile Ranks	Value
Rfree		0.249
Clashscore		3
Ramachandran outliers		0.3%
Sidechain outliers		8.7%
RSRZ outliers		2.5%

Worse   *Better*

▌ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2307 (2.94-2.90)
Clashscore	141614	2531 (2.94-2.90)
Ramachandran outliers	138981	2462 (2.94-2.90)
Sidechain outliers	138945	2464 (2.94-2.90)
RSRZ outliers	127900	2248 (2.94-2.90)

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 <=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	520	2% · 74% · 13% · 13%
1	B	520	2% · 74% · 11% · 13%

WORLDWIDE
wwPDB
PROTEIN DATA BANK

2 Entry composition ⓘ

There is only 1 type of molecule in this entry. The entry contains 7106 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 DUF1735 domain-containing protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	454	Total 3555	C 2271	N 575	O 696	S 13	0	0	0
1	B	454	Total 3551	C 2268	N 574	O 696	S 13	0	0	0

There are 68 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-12	MET	-	initiating methionine	UNP A0A6H0KVH8
A	-11	GLY	-	expression tag	UNP A0A6H0KVH8
A	-10	SER	-	expression tag	UNP A0A6H0KVH8
A	-9	SER	-	expression tag	UNP A0A6H0KVH8
A	-8	HIS	-	expression tag	UNP A0A6H0KVH8
A	-7	HIS	-	expression tag	UNP A0A6H0KVH8
A	-6	HIS	-	expression tag	UNP A0A6H0KVH8
A	-5	HIS	-	expression tag	UNP A0A6H0KVH8
A	-4	HIS	-	expression tag	UNP A0A6H0KVH8
A	-3	HIS	-	expression tag	UNP A0A6H0KVH8
A	-2	SER	-	expression tag	UNP A0A6H0KVH8
A	-1	SER	-	expression tag	UNP A0A6H0KVH8
A	0	GLY	-	expression tag	UNP A0A6H0KVH8
A	1	LEU	-	expression tag	UNP A0A6H0KVH8
A	2	VAL	-	expression tag	UNP A0A6H0KVH8
A	3	PRO	-	expression tag	UNP A0A6H0KVH8
A	4	ARG	-	expression tag	UNP A0A6H0KVH8
A	5	GLY	-	expression tag	UNP A0A6H0KVH8
A	6	SER	-	expression tag	UNP A0A6H0KVH8
A	7	HIS	-	expression tag	UNP A0A6H0KVH8
A	8	MET	-	expression tag	UNP A0A6H0KVH8
A	9	ALA	-	expression tag	UNP A0A6H0KVH8
A	10	SER	-	expression tag	UNP A0A6H0KVH8
A	11	MET	-	expression tag	UNP A0A6H0KVH8
A	12	THR	-	expression tag	UNP A0A6H0KVH8

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	13	GLY	-	expression tag	UNP A0A6H0KVH8
A	14	GLY	-	expression tag	UNP A0A6H0KVH8
A	15	GLN	-	expression tag	UNP A0A6H0KVH8
A	16	GLN	-	expression tag	UNP A0A6H0KVH8
A	17	MET	-	expression tag	UNP A0A6H0KVH8
A	18	GLY	-	expression tag	UNP A0A6H0KVH8
A	19	ARG	-	expression tag	UNP A0A6H0KVH8
A	20	GLY	-	expression tag	UNP A0A6H0KVH8
A	21	SER	-	expression tag	UNP A0A6H0KVH8
B	-12	MET	-	initiating methionine	UNP A0A6H0KVH8
B	-11	GLY	-	expression tag	UNP A0A6H0KVH8
B	-10	SER	-	expression tag	UNP A0A6H0KVH8
B	-9	SER	-	expression tag	UNP A0A6H0KVH8
B	-8	HIS	-	expression tag	UNP A0A6H0KVH8
B	-7	HIS	-	expression tag	UNP A0A6H0KVH8
B	-6	HIS	-	expression tag	UNP A0A6H0KVH8
B	-5	HIS	-	expression tag	UNP A0A6H0KVH8
B	-4	HIS	-	expression tag	UNP A0A6H0KVH8
B	-3	HIS	-	expression tag	UNP A0A6H0KVH8
B	-2	SER	-	expression tag	UNP A0A6H0KVH8
B	-1	SER	-	expression tag	UNP A0A6H0KVH8
B	0	GLY	-	expression tag	UNP A0A6H0KVH8
B	1	LEU	-	expression tag	UNP A0A6H0KVH8
B	2	VAL	-	expression tag	UNP A0A6H0KVH8
B	3	PRO	-	expression tag	UNP A0A6H0KVH8
B	4	ARG	-	expression tag	UNP A0A6H0KVH8
B	5	GLY	-	expression tag	UNP A0A6H0KVH8
B	6	SER	-	expression tag	UNP A0A6H0KVH8
B	7	HIS	-	expression tag	UNP A0A6H0KVH8
B	8	MET	-	expression tag	UNP A0A6H0KVH8
B	9	ALA	-	expression tag	UNP A0A6H0KVH8
B	10	SER	-	expression tag	UNP A0A6H0KVH8
B	11	MET	-	expression tag	UNP A0A6H0KVH8
B	12	THR	-	expression tag	UNP A0A6H0KVH8
B	13	GLY	-	expression tag	UNP A0A6H0KVH8
B	14	GLY	-	expression tag	UNP A0A6H0KVH8
B	15	GLN	-	expression tag	UNP A0A6H0KVH8
B	16	GLN	-	expression tag	UNP A0A6H0KVH8
B	17	MET	-	expression tag	UNP A0A6H0KVH8
B	18	GLY	-	expression tag	UNP A0A6H0KVH8
B	19	ARG	-	expression tag	UNP A0A6H0KVH8
B	20	GLY	-	expression tag	UNP A0A6H0KVH8

Continued from previous page...

| Chain	Residue	Modelled	Actual	Comment	Reference
B	21	SER	-	expression tag	UNP A0A6H0KVH8

3 Residue-property plots ⓘ

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: DUF1735 domain-containing protein
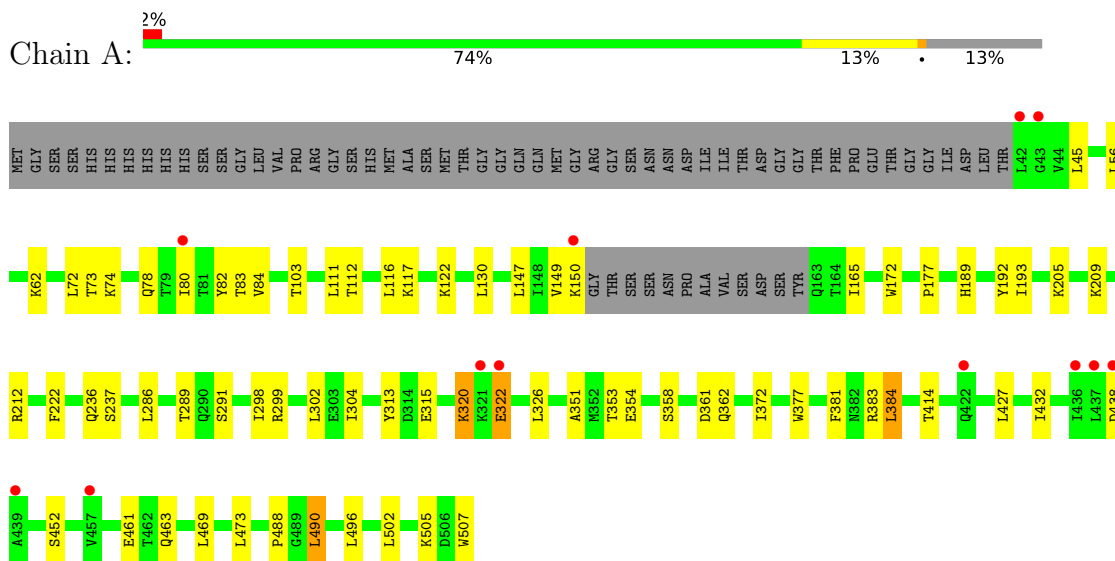
• Molecule 1: DUF1735 domain-containing protein
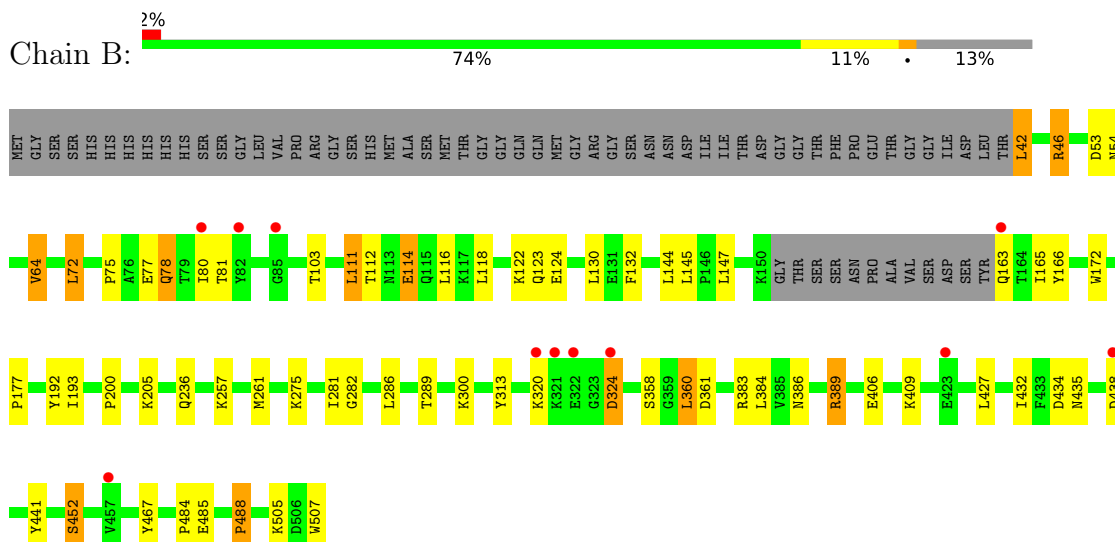
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	74.26Å 57.54Å 137.15Å 90.00° 91.77° 90.00°	Depositor
Resolution (Å)	37.11 – 2.91 37.11 – 2.91	Depositor EDS
% Data completeness (in resolution range)	97.3 (37.11-2.91) 97.3 (37.11-2.91)	Depositor EDS
R_{merge}	0.19	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.67 (at 2.90Å)	Xtriage
Refinement program	PHENIX (1.20.1_4487: ???)	Depositor
R, R_{free}	0.201 , 0.252 0.204 , 0.249	Depositor DCC
R_{free} test set	2003 reflections (7.93%)	wwPDB-VP
Wilson B-factor (Å^2)	49.6	Xtriage
Anisotropy	0.233	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 52.1	EDS
L-test for twinning[2]	$< |L| > = 0.52, < L^2 > = 0.35$	Xtriage
Estimated twinning fraction	0.016 for h,-k,-l	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	7106	wwPDB-VP
Average B, all atoms (Å^2)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.13% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $< |L| >$, $< L^2 >$ 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	0.27	0/3630	0.49	0/4932
1	B	0.27	0/3626	0.48	0/4928
All	All	0.27	0/7256	0.48	0/9860

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	212	ARG	Sidechain

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	3555	0	3480	23	0
1	B	3551	0	3469	26	0
All	All	7106	0	6949	49	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 3.

All (49) 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:B:192:TYR:HB2	1:B:452:SER:H	1.27	0.98
1:A:192:TYR:HB2	1:A:452:SER:H	1.43	0.82
1:B:193:ILE:HD13	1:B:200:PRO:HA	1.70	0.74
1:B:78:GLN:HG3	1:B:80:ILE:HD11	1.72	0.72
1:A:384:LEU:HD13	1:A:432:ILE:HA	1.72	0.70
1:B:275:LYS:HG2	1:B:313:TYR:HB3	1.83	0.59
1:A:192:TYR:HB2	1:A:452:SER:N	2.14	0.58
1:B:111:LEU:HD13	1:B:114:GLU:HG2	1.85	0.58
1:A:45:LEU:HD22	1:A:149:VAL:HG22	1.85	0.58
1:B:46:ARG:NH1	1:B:124:GLU:OE2	2.35	0.58
1:B:324:ASP:N	1:B:324:ASP:OD1	2.37	0.57
1:B:64:VAL:HG12	1:B:132:PHE:HB2	1.88	0.54
1:B:505:LYS:HE3	1:B:507:TRP:CE2	2.43	0.53
1:B:132:PHE:HE2	1:B:145:LEU:HD22	1.75	0.52
1:B:383:ARG:HD3	1:B:427:LEU:HD21	1.92	0.52
1:A:130:LEU:HD21	1:A:147:LEU:HD21	1.92	0.51
1:B:384:LEU:HD13	1:B:432:ILE:HA	1.93	0.50
1:A:189:HIS:HB3	1:A:222:PHE:HA	1.95	0.49
1:B:205:LYS:O	1:B:488:PRO:HD3	2.14	0.48
1:A:122:LYS:HB3	1:A:122:LYS:HE2	1.70	0.47
1:A:381:PHE:HA	1:A:414:THR:O	2.15	0.47
1:A:45:LEU:HD23	1:A:165:ILE:HG22	1.98	0.46
1:B:389:ARG:NH2	1:B:434:ASP:OD2	2.43	0.45
1:A:320:LYS:HD3	1:A:322:GLU:H	1.82	0.45
1:A:298:ILE:O	1:A:302:LEU:HG	2.17	0.44
1:A:315:GLU:HB2	1:A:353:THR:HG21	1.99	0.44
1:B:144:LEU:HD11	1:B:166:TYR:HB3	2.00	0.44
1:A:172:TRP:CE2	1:A:177:PRO:HG3	2.53	0.43
1:A:505:LYS:HE3	1:A:507:TRP:NE1	2.33	0.43
1:B:406:GLU:HB2	1:B:409:LYS:HG3	2.01	0.43
1:A:383:ARG:HG3	1:A:427:LEU:HD21	2.00	0.43
1:B:281:ILE:HD11	1:B:286:LEU:HD23	2.00	0.43
1:B:300:LYS:HE2	1:B:300:LYS:HB3	1.83	0.43
1:B:282:GLY:O	1:B:286:LEU:HG	2.19	0.42
1:B:54:ASN:N	1:B:54:ASN:OD1	2.52	0.42
1:A:192:TYR:HH	1:A:377:TRP:HZ2	1.65	0.42
1:B:360:LEU:HD13	1:B:360:LEU:HA	1.89	0.42

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:82:TYR:O	1:A:116:LEU:N	2.44	0.42
1:A:286:LEU:HD13	1:A:291:SER:HA	2.01	0.41
1:B:77:GLU:H	1:B:77:GLU:HG2	1.66	0.41
1:B:435:ASN:O	1:B:441:TYR:HB2	2.20	0.41
1:A:313:TYR:HA	1:A:351:ALA:HB3	2.02	0.41
1:A:304:ILE:HD13	1:A:304:ILE:HA	1.81	0.41
1:B:42:LEU:O	1:B:72:LEU:HD23	2.21	0.41
1:A:205:LYS:O	1:A:488:PRO:HD3	2.21	0.41
1:B:75:PRO:HA	1:B:123:GLN:HG2	2.02	0.41
1:B:172:TRP:CE2	1:B:177:PRO:HG3	2.56	0.41
1:A:490:LEU:HD13	1:A:490:LEU:HA	1.87	0.40
1:A:56:LEU:HD23	1:A:209:LYS:HD2	2.04	0.40

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	450/520 (86%)	434 (96%)	16 (4%)	0	100	100
1	B	450/520 (86%)	438 (97%)	9 (2%)	3 (1%)	22	53
All	All	900/1040 (86%)	872 (97%)	25 (3%)	3 (0%)	41	70

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	452	SER
1	B	484	PRO
1	B	488	PRO

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	386/440 (88%)	351 (91%)	35 (9%)	9	27
1	B	385/440 (88%)	353 (92%)	32 (8%)	11	30
All	All	771/880 (88%)	704 (91%)	67 (9%)	10	29

All (67) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	62	LYS
1	A	72	LEU
1	A	73	THR
1	A	74	LYS
1	A	78	GLN
1	A	80	ILE
1	A	83	THR
1	A	84	VAL
1	A	103	THR
1	A	111	LEU
1	A	112	THR
1	A	117	LYS
1	A	150	LYS
1	A	193	ILE
1	A	236	GLN
1	A	237	SER
1	A	289	THR
1	A	299	ARG
1	A	320	LYS
1	A	322	GLU
1	A	326	LEU
1	A	354	GLU
1	A	358	SER
1	A	361	ASP
1	A	362	GLN
1	A	372	ILE
1	A	384	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	438	ASP
1	A	461	GLU
1	A	463	GLN
1	A	469	LEU
1	A	473	LEU
1	A	490	LEU
1	A	496	LEU
1	A	502	LEU
1	B	42	LEU
1	B	46	ARG
1	B	53	ASP
1	B	64	VAL
1	B	72	LEU
1	B	78	GLN
1	B	81	THR
1	B	103	THR
1	B	111	LEU
1	B	112	THR
1	B	114	GLU
1	B	116	LEU
1	B	118	LEU
1	B	122	LYS
1	B	130	LEU
1	B	147	LEU
1	B	163	GLN
1	B	165	ILE
1	B	236	GLN
1	B	257	LYS
1	B	261	MET
1	B	289	THR
1	B	320	LYS
1	B	324	ASP
1	B	358	SER
1	B	360	LEU
1	B	361	ASP
1	B	386	ASN
1	B	389	ARG
1	B	438	ASP
1	B	467	TYR
1	B	485	GLU

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	137	ASN
1	B	386	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 ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

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(Å²)	Q<0.9
1	A	454/520 (87%)	-0.12	12 (2%)	56	53	23, 41, 65, 91	0
1	B	454/520 (87%)	-0.13	11 (2%)	59	57	28, 41, 64, 90	0
All	All	908/1040 (87%)	-0.12	23 (2%)	57	56	23, 41, 65, 91	0

All (23) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	43	GLY	4.4
1	A	438	ASP	3.5
1	B	324	ASP	3.3
1	A	322	GLU	3.2
1	B	82	TYR	2.9
1	A	42	LEU	2.7
1	B	80	ILE	2.6
1	A	150	LYS	2.6
1	B	322	GLU	2.6
1	B	320	LYS	2.6
1	A	80	ILE	2.6
1	B	163	GLN	2.5
1	B	438	ASP	2.5
1	A	439	ALA	2.4
1	A	321	LYS	2.4
1	B	321	LYS	2.3
1	B	423	GLU	2.3
1	A	437	LEU	2.2
1	A	457	VAL	2.1
1	B	457	VAL	2.0
1	A	436	ILE	2.0
1	B	85	GLY	2.0
1	A	422	GLN	2.0

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 monosaccharides 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.