wwPDB X-ray Structure Validation Summary Report (i)

Jun 12, 2024 – 04:37 PM EDT

PDB ID	:	1AUS
Title	:	ACTIVATED UNLIGANDED SPINACH RUBISCO
Authors	:	Taylor, T.C.; Andersson, I.
Deposited on	:	1995-06-21
Resolution	:	2.20 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 (i) 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 (1)) were used in the production of this report:

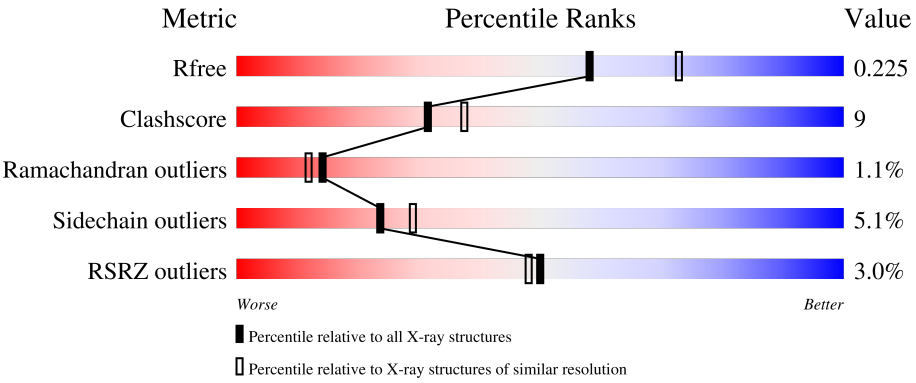
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.20 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	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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	L	475	2% 75% 15% • 8%
1	M	475	3% 76% 15% • 8%
1	N	475	3% 75% 16% • 8%
1	O	475	2% 75% 15% • 8%
2	S	123	7% 75% 20% 5%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	T	123	6% 76% 21%
2	U	123	% 75% 22%
2	V	123	4% 76% 21%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 18149 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 RIBULOSE BISPHOSPHATE CARBOXYLASE/OXYGENASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	L	439	Total	C	N	O	S	0	0	0
			3440	2180	608	635	17			
1	M	439	Total	C	N	O	S	0	0	0
			3440	2180	608	635	17			
1	N	439	Total	C	N	O	S	0	0	0
			3440	2180	608	635	17			
1	O	439	Total	C	N	O	S	0	0	0
			3440	2180	608	635	17			

- Molecule 2 is a protein called RIBULOSE BISPHOSPHATE CARBOXYLASE/OXYGENASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	S	123	Total	C	N	O	S	0	0	0
			1033	673	167	186	7			
2	T	123	Total	C	N	O	S	0	0	0
			1033	673	167	186	7			
2	U	123	Total	C	N	O	S	0	0	0
			1033	673	167	186	7			
2	V	123	Total	C	N	O	S	0	0	0
			1033	673	167	186	7			

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
S	2	GLN	LYS	CONFLICT	UNP Q43832
S	6	ILE	THR	CONFLICT	UNP Q43832
S	7	LEU	GLN	CONFLICT	UNP Q43832
S	9	LEU	MET	CONFLICT	UNP Q43832
S	11	LYS	ARG	CONFLICT	UNP Q43832
S	109	GLU	GLN	CONFLICT	UNP Q43832
S	113	ILE	VAL	CONFLICT	UNP Q43832

Continued on next page...

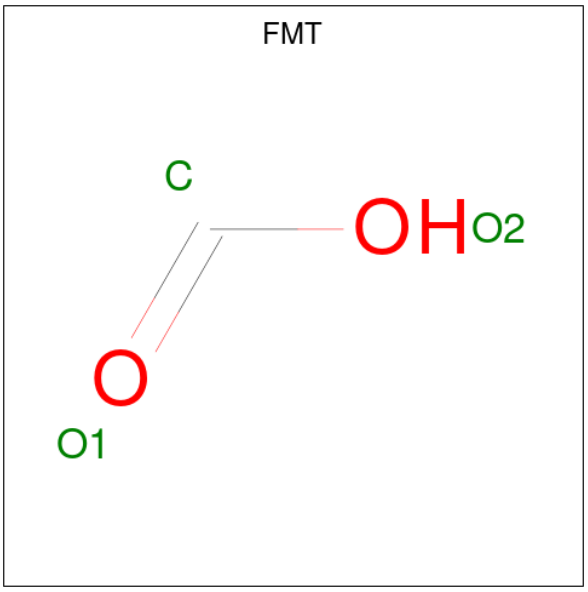
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
T	2	GLN	LYS	CONFLICT	UNP Q43832
T	6	ILE	THR	CONFLICT	UNP Q43832
T	7	LEU	GLN	CONFLICT	UNP Q43832
T	9	LEU	MET	CONFLICT	UNP Q43832
T	11	LYS	ARG	CONFLICT	UNP Q43832
T	109	GLU	GLN	CONFLICT	UNP Q43832
T	113	ILE	VAL	CONFLICT	UNP Q43832
U	2	GLN	LYS	CONFLICT	UNP Q43832
U	6	ILE	THR	CONFLICT	UNP Q43832
U	7	LEU	GLN	CONFLICT	UNP Q43832
U	9	LEU	MET	CONFLICT	UNP Q43832
U	11	LYS	ARG	CONFLICT	UNP Q43832
U	109	GLU	GLN	CONFLICT	UNP Q43832
U	113	ILE	VAL	CONFLICT	UNP Q43832
V	2	GLN	LYS	CONFLICT	UNP Q43832
V	6	ILE	THR	CONFLICT	UNP Q43832
V	7	LEU	GLN	CONFLICT	UNP Q43832
V	9	LEU	MET	CONFLICT	UNP Q43832
V	11	LYS	ARG	CONFLICT	UNP Q43832
V	109	GLU	GLN	CONFLICT	UNP Q43832
V	113	ILE	VAL	CONFLICT	UNP Q43832

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	L	1	Total Mg 1 1	0	0
3	M	1	Total Mg 1 1	0	0
3	N	1	Total Mg 1 1	0	0
3	O	1	Total Mg 1 1	0	0

- Molecule 4 is FORMIC ACID (three-letter code: FMT) (formula: CH_2O_2).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	L	1	Total C O 3 1 2	0	0
4	M	1	Total C O 3 1 2	0	0
4	N	1	Total C O 3 1 2	0	0
4	O	1	Total C O 3 1 2	0	0

- Molecule 5 is water.

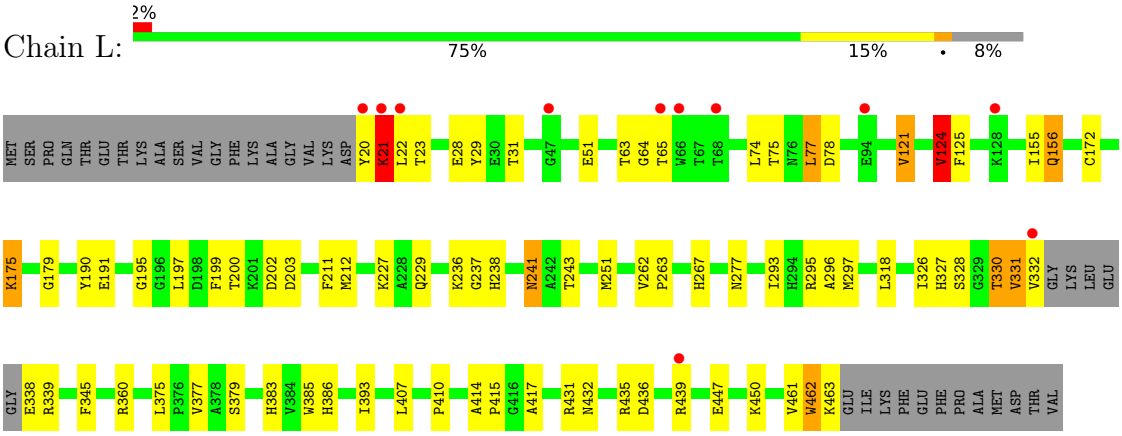
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	L	187	Total O 187 187	0	0
5	S	39	Total O 39 39	0	0
5	M	4	Total O 4 4	0	0
5	T	1	Total O 1 1	0	0
5	O	3	Total O 3 3	0	0
5	V	7	Total O 7 7	0	0

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.

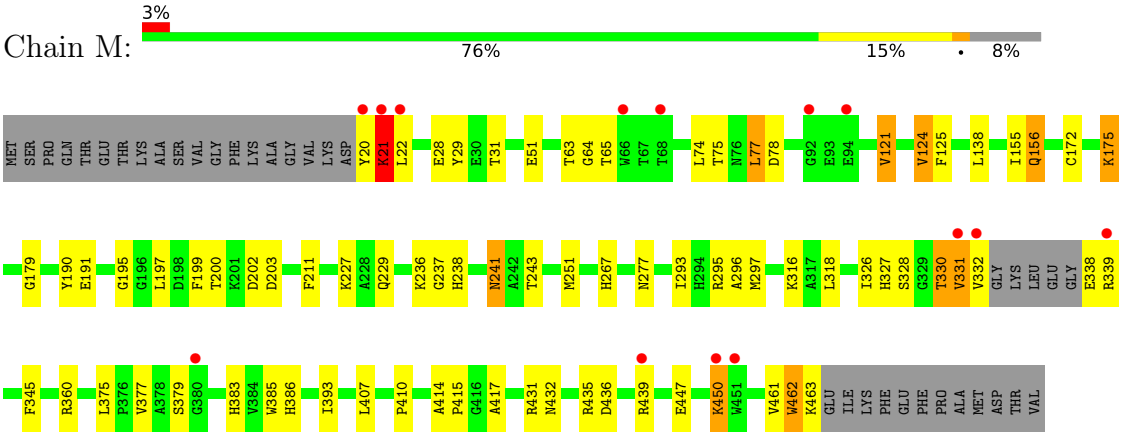
• Molecule 1: RIBULOSE BISPHOSPHATE CARBOXYLASE/OXYGENASE

Chain L:



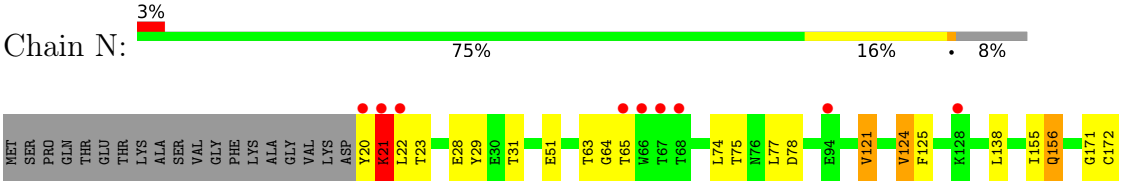
• Molecule 1: RIBULOSE BISPHOSPHATE CARBOXYLASE/OXYGENASE

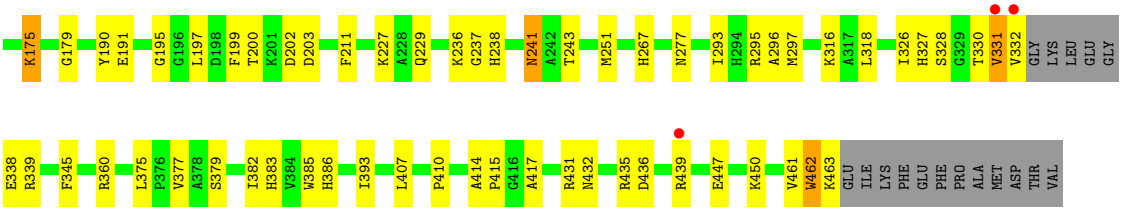
Chain M:



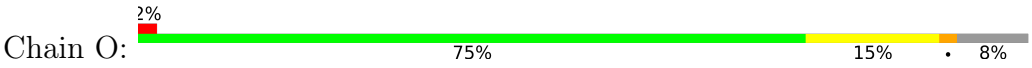
• Molecule 1: RIBULOSE BISPHOSPHATE CARBOXYLASE/OXYGENASE

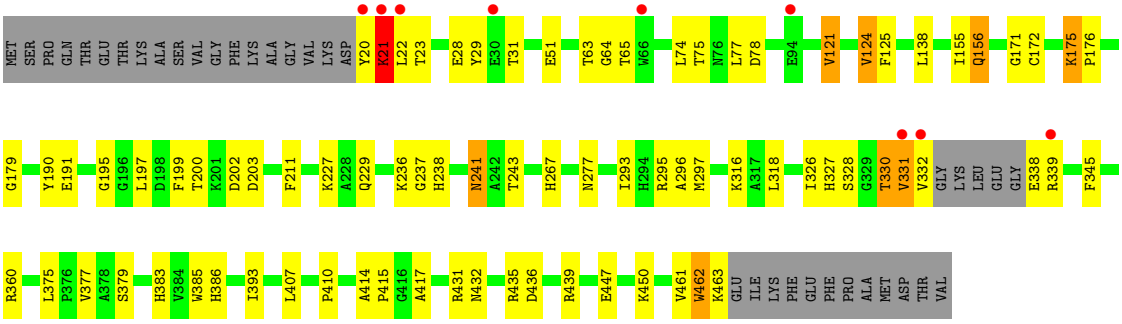
Chain N:



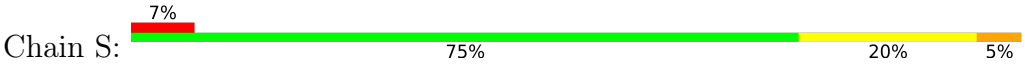


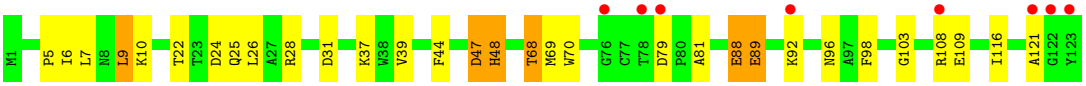
• Molecule 1: RIBULOSE BISPHOSPHATE CARBOXYLASE/OXYGENASE



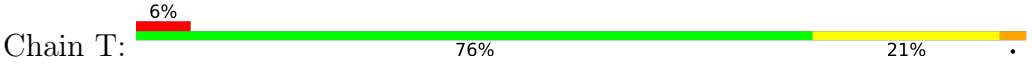


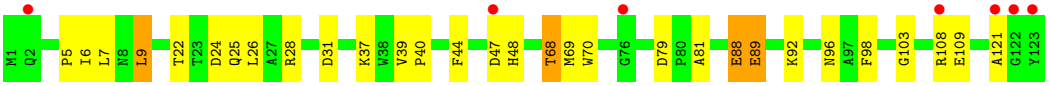
• Molecule 2: RIBULOSE BISPHOSPHATE CARBOXYLASE/OXYGENASE



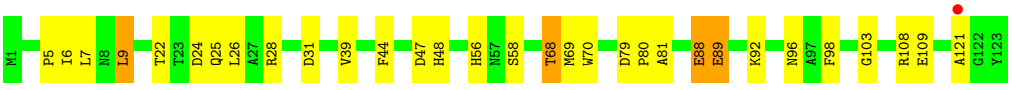


• Molecule 2: RIBULOSE BISPHOSPHATE CARBOXYLASE/OXYGENASE

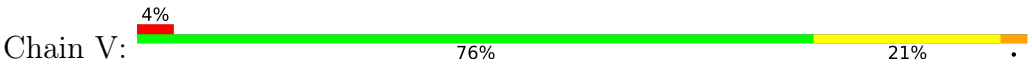


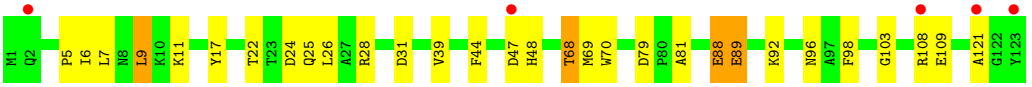


• Molecule 2: RIBULOSE BISPHOSPHATE CARBOXYLASE/OXYGENASE



• Molecule 2: RIBULOSE BISPHOSPHATE CARBOXYLASE/OXYGENASE





4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	157.40Å 158.70Å 203.00Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	7.00 – 2.20 19.89 – 2.10	Depositor EDS
% Data completeness (in resolution range)	94.8 (7.00-2.20) 89.0 (19.89-2.10)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.68 (at 2.09Å)	Xtriage
Refinement program	X-PLOR	Depositor
R, R_{free}	0.217 , 0.239 0.220 , 0.225	Depositor DCC
R_{free} test set	6637 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å^2)	17.7	Xtriage
Anisotropy	0.102	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 52.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.016 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	18149	wwPDB-VP
Average B, all atoms (Å^2)	15.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 48.57 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 8.4383e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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

Bond lengths and bond angles in the following residue types are not validated in this section: FMT, 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	L	0.67	0/3525	0.81	2/4785 (0.0%)
1	M	0.67	0/3525	0.81	2/4785 (0.0%)
1	N	0.67	0/3525	0.81	2/4785 (0.0%)
1	O	0.67	0/3525	0.81	2/4785 (0.0%)
2	S	0.74	0/1068	0.81	0/1453
2	T	0.74	0/1068	0.81	0/1453
2	U	0.74	0/1068	0.81	0/1453
2	V	0.74	0/1068	0.81	0/1453
All	All	0.69	0/18372	0.81	8/24952 (0.0%)

There are no bond length outliers.

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	O	462	TRP	N-CA-C	6.94	129.73	111.00
1	L	462	TRP	N-CA-C	6.94	129.73	111.00
1	M	462	TRP	N-CA-C	6.94	129.73	111.00
1	N	462	TRP	N-CA-C	6.93	129.72	111.00
1	N	124	VAL	CB-CA-C	-5.86	100.27	111.40

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	L	3440	0	3353	65	2
1	M	3440	0	3353	58	3
1	N	3440	0	3353	63	2
1	O	3440	0	3353	63	2
2	S	1033	0	990	28	2
2	T	1033	0	990	25	2
2	U	1033	0	990	26	2
2	V	1033	0	990	26	2
3	L	1	0	0	0	0
3	M	1	0	0	0	0
3	N	1	0	0	0	0
3	O	1	0	0	0	0
4	L	3	0	0	1	0
4	M	3	0	0	1	0
4	N	3	0	0	1	0
4	O	3	0	0	1	0
5	L	187	0	0	5	0
5	M	4	0	0	0	0
5	O	3	0	0	0	0
5	S	39	0	0	1	0
5	T	1	0	0	0	0
5	V	7	0	0	0	0
All	All	18149	0	17372	313	9

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 9.

The worst 5 of 313 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:L:179:GLY:O	2:V:109:GLU:HG2	1.76	0.85
1:N:267:HIS:HD2	1:N:277:ASN:HD22	1.25	0.85
1:O:267:HIS:HD2	1:O:277:ASN:HD22	1.25	0.84
1:M:267:HIS:HD2	1:M:277:ASN:HD22	1.25	0.83
1:L:267:HIS:HD2	1:L:277:ASN:HD22	1.25	0.82

The worst 5 of 9 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)
1:L:74:LEU:O	2:T:109:GLU:OE1[3_555]	1.37	0.83

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:S:109:GLU:OE1	1:M:74:LEU:O[3_555]	1.37	0.83
1:N:74:LEU:O	2:V:109:GLU:OE1[3_555]	1.40	0.80
2:U:109:GLU:OE1	1:O:74:LEU:O[3_555]	1.40	0.80
2:S:109:GLU:OE1	1:M:74:LEU:C[3_555]	1.86	0.34

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	L	435/475 (92%)	416 (96%)	15 (3%)	4 (1%)	17	16
1	M	435/475 (92%)	416 (96%)	15 (3%)	4 (1%)	17	16
1	N	435/475 (92%)	416 (96%)	15 (3%)	4 (1%)	17	16
1	O	435/475 (92%)	416 (96%)	15 (3%)	4 (1%)	17	16
2	S	121/123 (98%)	112 (93%)	7 (6%)	2 (2%)	9	6
2	T	121/123 (98%)	112 (93%)	7 (6%)	2 (2%)	9	6
2	U	121/123 (98%)	112 (93%)	7 (6%)	2 (2%)	9	6
2	V	121/123 (98%)	112 (93%)	7 (6%)	2 (2%)	9	6
All	All	2224/2392 (93%)	2112 (95%)	88 (4%)	24 (1%)	14	12

5 of 24 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	L	63	THR
1	L	331	VAL
2	S	47	ASP
1	M	63	THR
1	M	331	VAL

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	L	357/386 (92%)	340 (95%)	17 (5%)	25	32
1	M	357/386 (92%)	340 (95%)	17 (5%)	25	32
1	N	357/386 (92%)	341 (96%)	16 (4%)	27	34
1	O	357/386 (92%)	340 (95%)	17 (5%)	25	32
2	S	112/112 (100%)	105 (94%)	7 (6%)	18	20
2	T	112/112 (100%)	105 (94%)	7 (6%)	18	20
2	U	112/112 (100%)	105 (94%)	7 (6%)	18	20
2	V	112/112 (100%)	105 (94%)	7 (6%)	18	20
All	All	1876/1992 (94%)	1781 (95%)	95 (5%)	24	29

5 of 95 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	N	172	CYS
1	O	21	LYS
1	N	241	ASN
2	U	26	LEU
1	O	77	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 65 such sidechains are listed below:

Mol	Chain	Res	Type
1	O	267	HIS
1	O	304	GLN
1	M	241	ASN
1	M	238	HIS
1	O	386	HIS

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 8 ligands modelled in this entry, 4 are monoatomic - leaving 4 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
4	FMT	N	477	1,3	2,2,2	0.95	0	1,1,1	0.88	0
4	FMT	O	477	1,3	2,2,2	0.96	0	1,1,1	0.88	0
4	FMT	L	477	1,3	2,2,2	0.95	0	1,1,1	0.88	0
4	FMT	M	477	1,3	2,2,2	0.94	0	1,1,1	0.88	0

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.

4 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	N	477	FMT	1	0
4	O	477	FMT	1	0
4	L	477	FMT	1	0
4	M	477	FMT	1	0

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, 95^{th} 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	L	439/475 (92%)	-0.34	11 (2%)	57	55	4, 11, 30, 39	0
1	M	439/475 (92%)	-0.30	14 (3%)	47	45	4, 11, 30, 39	0
1	N	439/475 (92%)	-0.35	12 (2%)	54	52	4, 11, 30, 39	0
1	O	439/475 (92%)	-0.36	9 (2%)	63	61	4, 11, 30, 39	0
2	S	123/123 (100%)	0.03	8 (6%)	18	17	6, 20, 32, 38	0
2	T	123/123 (100%)	0.22	7 (5%)	23	22	6, 20, 32, 38	0
2	U	123/123 (100%)	-0.04	1 (0%)	86	85	6, 20, 32, 38	0
2	V	123/123 (100%)	-0.08	5 (4%)	37	35	6, 20, 32, 38	0
All	All	2248/2392 (93%)	-0.26	67 (2%)	50	48	4, 14, 31, 39	0

The worst 5 of 67 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	M	332	VAL	7.3
1	M	22	LEU	7.2
1	L	332	VAL	6.9
1	O	22	LEU	6.6
1	N	332	VAL	6.5

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 (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, 95^{th} 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(Å^2)	Q<0.9
4	FMT	L	477	3/3	0.92	0.12	15,15,15,16	0
3	MG	N	476	1/1	0.94	0.07	17,17,17,17	0
3	MG	M	476	1/1	0.94	0.07	17,17,17,17	0
3	MG	O	476	1/1	0.95	0.08	17,17,17,17	0
3	MG	L	476	1/1	0.95	0.06	17,17,17,17	0
4	FMT	M	477	3/3	0.95	0.16	15,15,15,16	0
4	FMT	N	477	3/3	0.97	0.11	15,15,15,16	0
4	FMT	O	477	3/3	0.98	0.10	15,15,15,16	0

6.5 Other polymers (i)

There are no such residues in this entry.