wwPDB X-ray Structure Validation Summary Report (i)

May 15, 2020 – 08:17 pm BST

PDB ID	:	5OT4
Title	:	Structure of the Legionella pneumophila effector RidL (1-866)
Authors	:	Romano-Moreno, M.; Rojas, A.L.; Lucas, M.; Isupov, M.N.; Hierro, A.
Deposited on	:	2017-08-20
Resolution	:	3.00 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 following versions of software and data (see references (1)) were used in the production of this report:

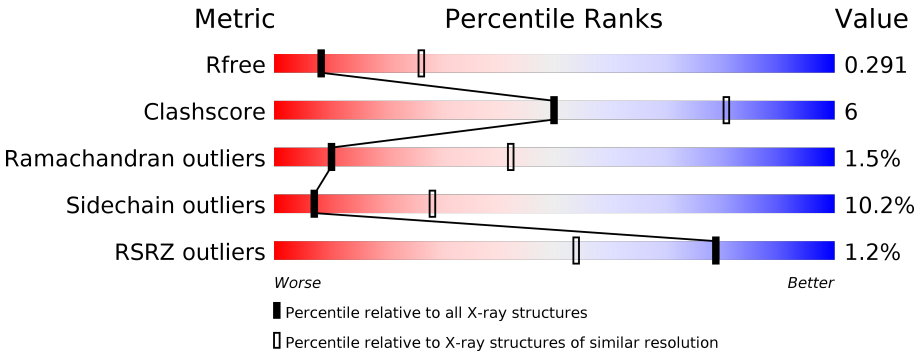
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 3.00 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	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)

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 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	901	% 76% 16% • 6%
1	B	901	76% 16% • 7%
1	C	901	3% 64% 19% • 13%
1	D	901	74% 16% • 7%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 26232 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.

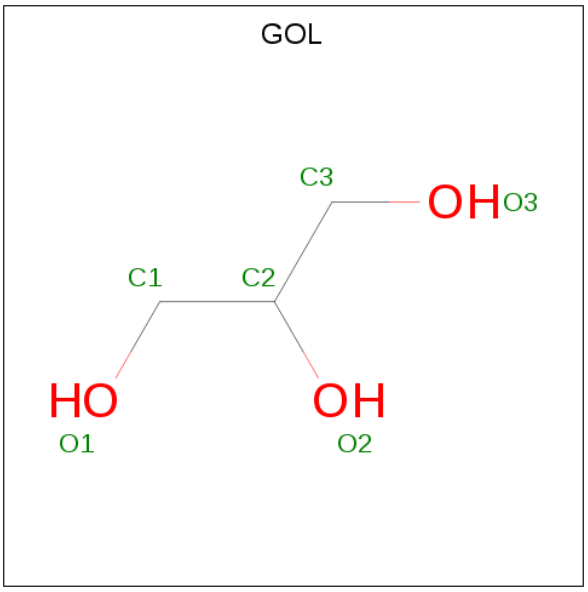
- Molecule 1 is a protein called Interaptin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	846	Total	C	N	O	S	0	0	0
			6687	4198	1164	1312	13			
1	B	842	Total	C	N	O	S	0	0	0
			6673	4191	1162	1306	14			
1	C	787	Total	C	N	O	S	0	0	0
			6238	3915	1084	1225	14			
1	D	836	Total	C	N	O	S	0	0	0
			6628	4158	1151	1305	14			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	SER	-	expression tag	UNP G8UZ99
B	0	SER	-	expression tag	UNP G8UZ99
C	0	SER	-	expression tag	UNP G8UZ99
D	0	SER	-	expression tag	UNP G8UZ99

- Molecule 2 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



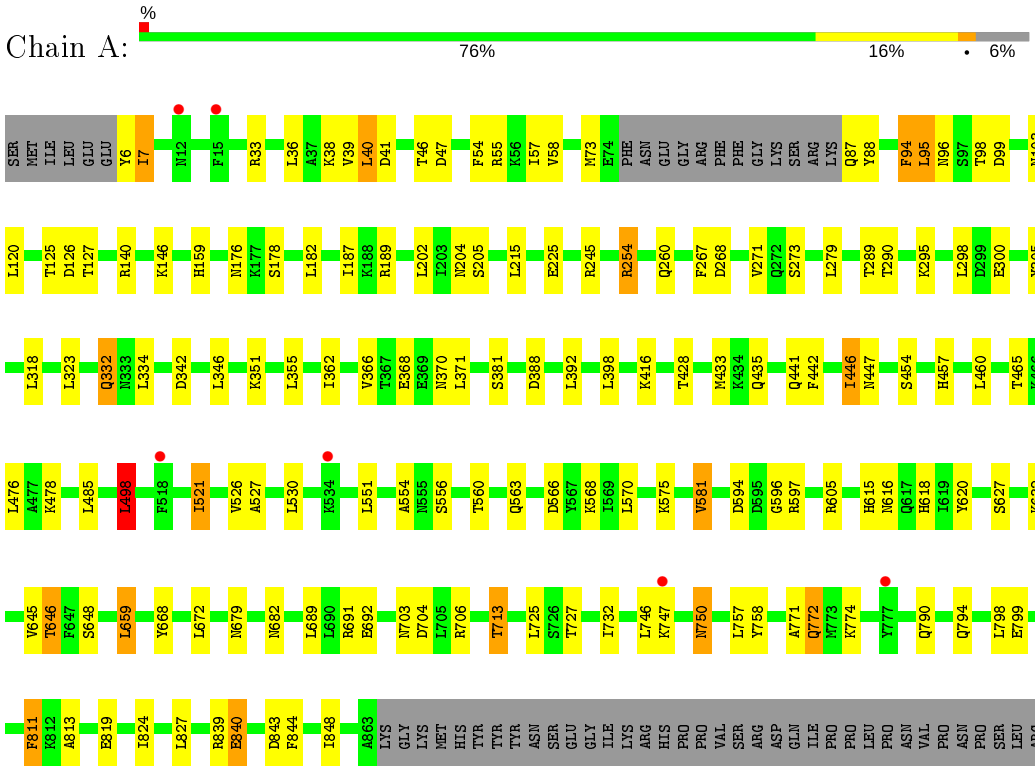
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
2	B	1	6	3	3	0	0

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.

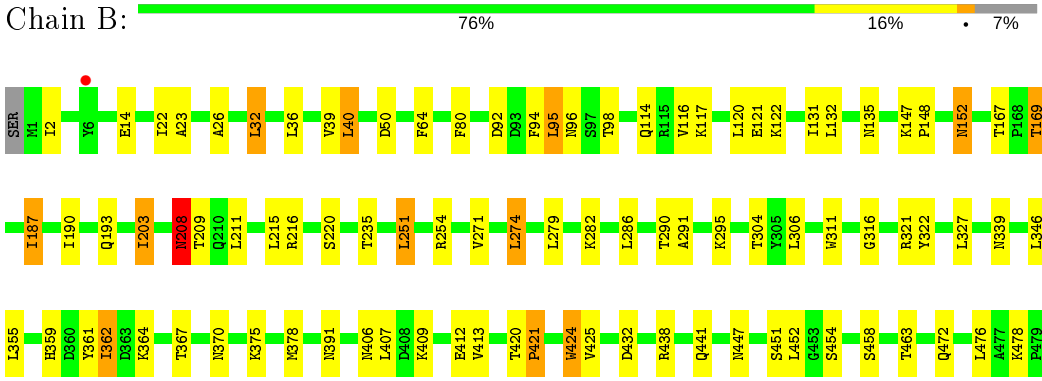
- Molecule 1: Interaptin

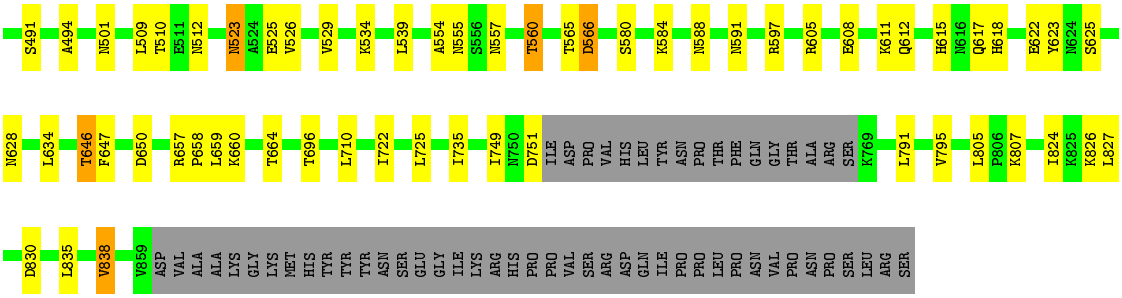
Chain A:



- Molecule 1: Interaptin

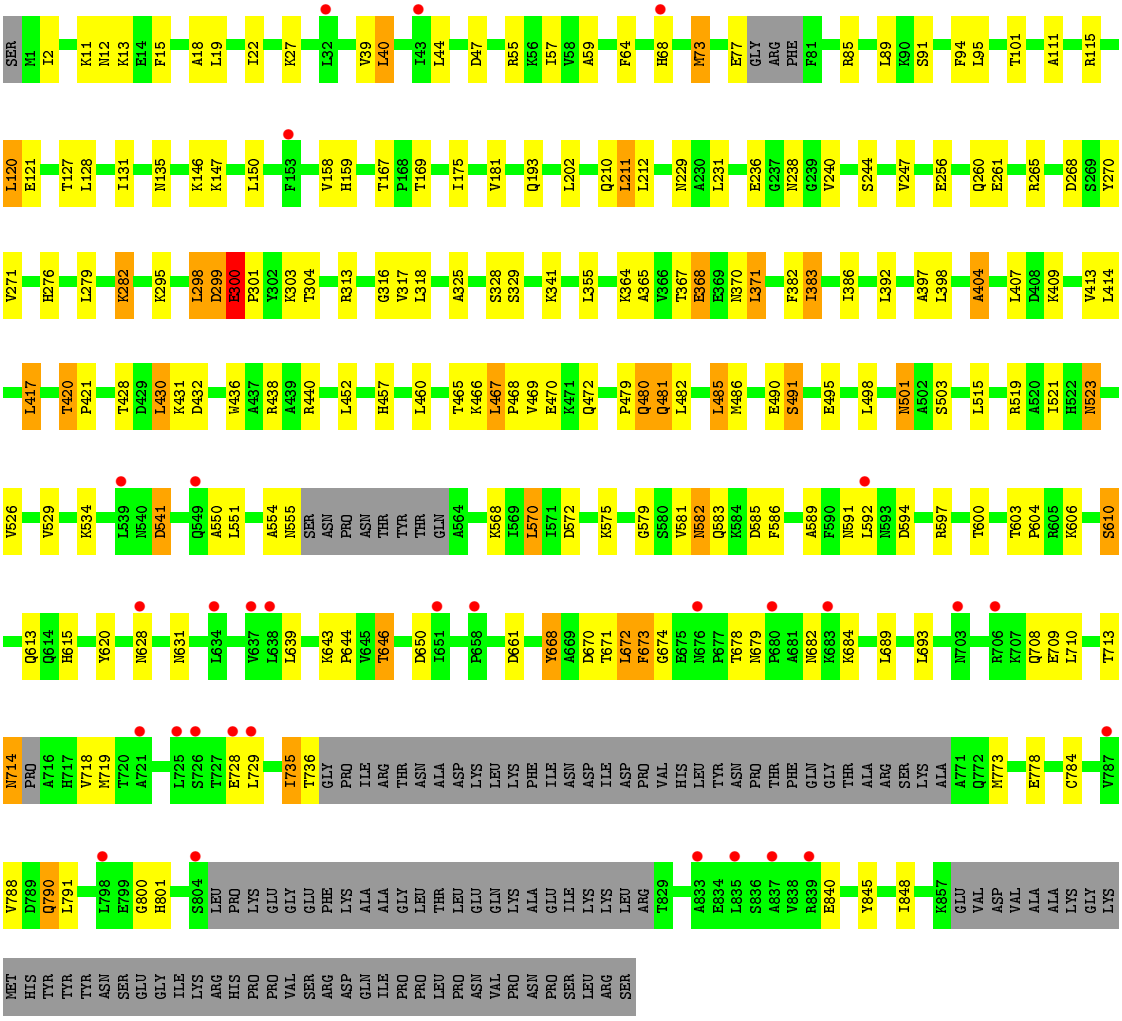
Chain B:





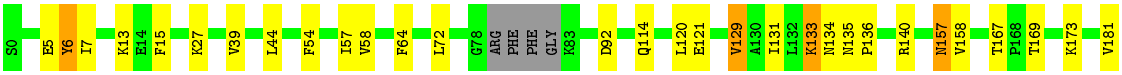
• Molecule 1: Interaptin

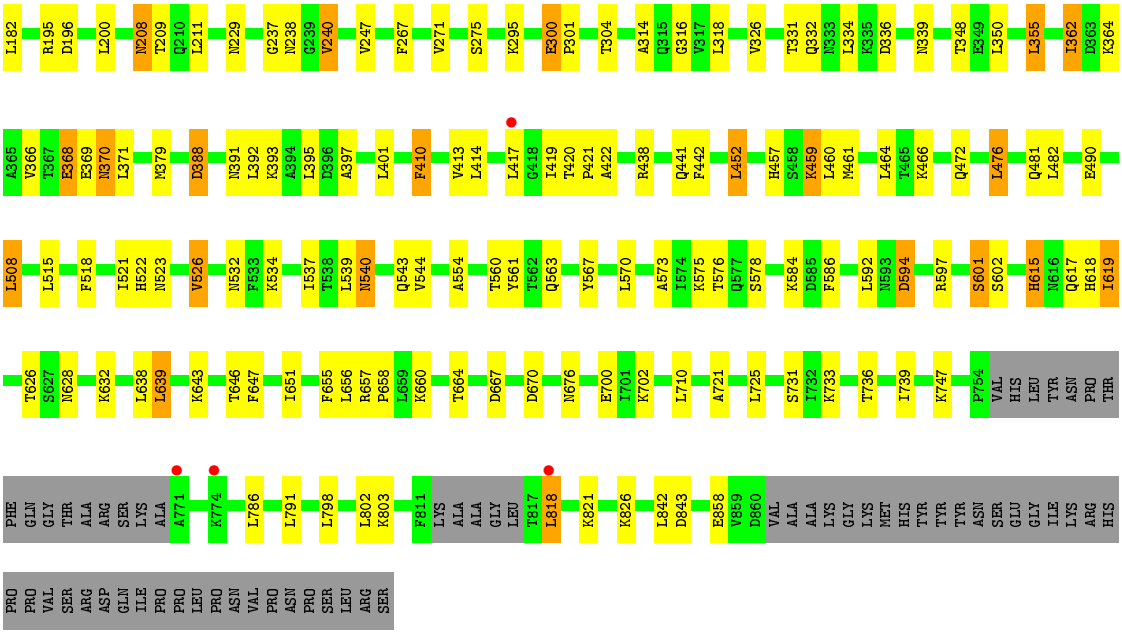
Chain C:



• Molecule 1: Interaptin

Chain D:




L182 R195 D196 L200 N208 T209 Q210 L211 N229 G237 N238 G239 V240 V247 F267 V271 S275 K295 E300 P301 T304 A314 Q315 G316 V317 L318 V326 T331 Q332 N333 L334 K335 D336 N339 T348 E349 L350 L385 I362 D363 K364
A365 V366 T367 E368 E369 N370 L371 M379 D388 N391 L392 K393 A394 L395 D396 A397 L401 F410 V413 L414 L417 G418 I419 T420 P421 A422 R438 Q441 F442 L452 H457 S458 K459 L460 M461 L464 T465 K466 Q472 L476 Q481 L482 E490
L508 L515 F518 I521 H522 N523 V526 N532 F533 K534 I537 T538 L539 N540 Q543 V544 A554 T560 Y561 T562 Q563 Y567 L570 A573 I574 K575 T576 Q577 S578 K584 D585 F586 L592 N593 D594 R597 S601 S602 H615 N616 Q617 H618 I619
T626 S627 N628 K632 L638 L639 K643 T646 F647 I651 F655 L656 R657 P658 L659 K660 T664 D667 D670 N676 E700 I701 K702 L710 A721 L725 S731 I732 K733 T736 I739 K747 P754 VAL HIS LEU TYR ASN PRO THR
PHE GLN GLY THR ALA ARG SER LYS ALA A771 K774 L786 L791 L798 L802 K803 F811 LYS ALA ALA GLY LEU T817 L818 K821 K826 L842 D843 E858 V869 D880 VAL ALA ALA LYS GLY LYS MET HIS TYR TYR TYR ASN SER GLU GLY ILE LYS ARG HIS
PRO PRO VAL SER ARG ASP GLN ILE PRO PRO LEU PRO ASN VAL PRO ASN PRO SER LEU ARG SER

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	242.65Å 230.91Å 86.73Å 90.00° 90.40° 90.00°	Depositor
Resolution (Å)	167.27 – 3.00 49.55 – 3.00	Depositor EDS
% Data completeness (in resolution range)	99.8 (167.27-3.00) 99.8 (49.55-3.00)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.49 (at 3.01Å)	Xtriage
Refinement program	REFMAC 5.8.0158	Depositor
R, R_{free}	0.229 , 0.295 0.230 , 0.291	Depositor DCC
R_{free} test set	4953 reflections (5.22%)	wwPDB-VP
Wilson B-factor (Å^2)	98.4	Xtriage
Anisotropy	0.043	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.26 , 59.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.000 for k,h,-l 0.000 for -k,-h,-l 0.025 for -h,-k,l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	26232	wwPDB-VP
Average B, all atoms (Å^2)	125.0	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 39.85 % 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 2.9839e-04. 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: GOL

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	0.52	0/6793	0.74	0/9163
1	B	0.52	0/6777	0.74	0/9130
1	C	0.59	0/6330	0.78	0/8523
1	D	0.53	0/6729	0.75	0/9067
All	All	0.54	0/26629	0.75	0/35883

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	6687	0	6729	64	0
1	B	6673	0	6726	71	0
1	C	6238	0	6270	93	0
1	D	6628	0	6666	79	0
2	B	6	0	8	0	0
All	All	26232	0	26399	307	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 6.

The worst 5 of 307 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:D:229:ASN:HD21	1:D:238:ASN:HD22	1.19	0.88
1:C:365:ALA:HA	1:C:370:ASN:HD22	1.48	0.79
1:C:212:LEU:HD23	1:C:231:LEU:HD11	1.65	0.78
1:C:229:ASN:HD21	1:C:238:ASN:HD21	1.33	0.76
1:B:116:VAL:HG21	1:B:190:ILE:HD13	1.70	0.73

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	842/901 (94%)	757 (90%)	74 (9%)	11 (1%)	12	45
1	B	838/901 (93%)	780 (93%)	50 (6%)	8 (1%)	15	53
1	C	775/901 (86%)	675 (87%)	78 (10%)	22 (3%)	5	25
1	D	828/901 (92%)	759 (92%)	60 (7%)	9 (1%)	14	50
All	All	3283/3604 (91%)	2971 (90%)	262 (8%)	50 (2%)	10	42

5 of 50 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	368	GLU
1	B	50	ASP
1	B	628	ASN
1	C	12	ASN
1	C	300	GLU

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	728/779 (94%)	656 (90%)	72 (10%)	8 30
1	B	726/779 (93%)	661 (91%)	65 (9%)	9 35
1	C	680/779 (87%)	594 (87%)	86 (13%)	4 20
1	D	725/779 (93%)	656 (90%)	69 (10%)	8 32
All	All	2859/3116 (92%)	2567 (90%)	292 (10%)	7 28

5 of 292 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	751	ASP
1	C	341	LYS
1	D	632	LYS
1	B	830	ASP
1	C	181	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 71 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	577	GLN
1	C	109	GLN
1	D	543	GLN
1	B	588	ASN
1	B	631	ASN

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 carbohydrates in this entry.

5.6 Ligand geometry (i)

1 ligand is modelled in this entry.

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	$\#\|Z\| > 2$
2	GOL	B	1001	-	5,5,5	0.38	0	5,5,5	0.28	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	GOL	B	1001	-	-	0/4/4/4	-

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.

No monomer is involved in short contacts.

5.7 Other polymers (i)

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, 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	A	846/901 (93%)	-0.32	6 (0%) 87 69	80, 116, 158, 196	0
1	B	842/901 (93%)	-0.32	1 (0%) 95 89	78, 111, 156, 210	0
1	C	787/901 (87%)	-0.09	30 (3%) 40 16	80, 141, 193, 219	0
1	D	836/901 (92%)	-0.31	4 (0%) 91 75	69, 127, 174, 204	0
All	All	3311/3604 (91%)	-0.26	41 (1%) 79 54	69, 122, 177, 219	0

The worst 5 of 41 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	771	ALA	5.2
1	C	68	HIS	3.7
1	C	726	SER	3.4
1	C	638	LEU	3.4
1	A	15	PHE	3.4

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

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
2	GOL	B	1001	6/6	0.82	0.15	155,167,170,171	0

6.5 Other polymers (i)

There are no such residues in this entry.