wwPDB X-ray Structure Validation Summary Report (i)

Mar 10, 2024 – 07:11 AM EDT

PDB ID	:	4IGL
Title	:	Structure of the RHS-repeat containing BC component of the secreted ABC toxin complex from Yersinia entomophaga
Authors	:	Busby, J.N.; Panjikar, S.; Landsberg, M.J.; Hurst, M.R.H.; Lott, J.S.
Deposited on	:	2012-12-17
Resolution	:	2.49 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 (i)) were used in the production of this report:

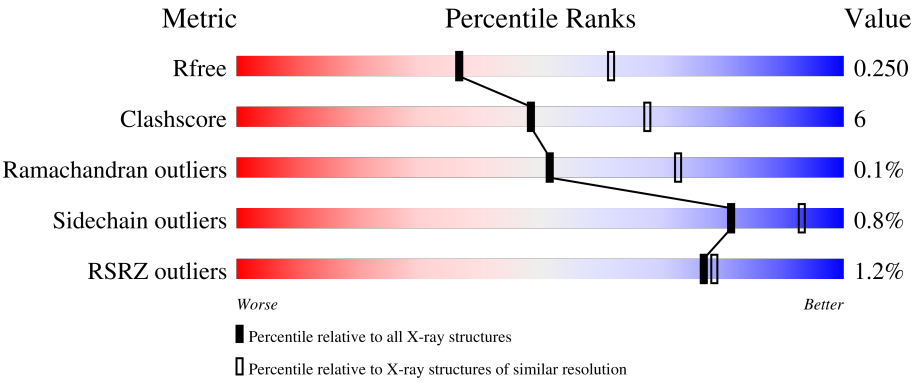
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.49 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	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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	1488	83% 14%
1	C	1488	83% 15%
2	B	690	83% 13%
2	D	690	80% 17%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 35548 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 YenB.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1457	Total 11431	C 7242	N 1963	O 2189	S 37	0	3	0
1	C	1449	Total 11370	C 7196	N 1948	O 2189	S 37	0	4	0

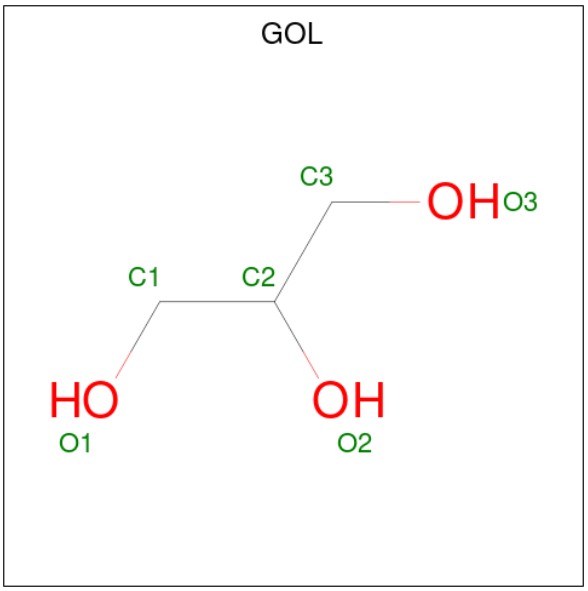
There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	GLY	-	expression tag	UNP B6A880
C	0	GLY	-	expression tag	UNP B6A880

- Molecule 2 is a protein called YenC2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	666	Total 5100	C 3181	N 886	O 1022	S 11	0	0	0
2	D	668	Total 5126	C 3201	N 891	O 1022	S 12	0	1	0

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 6 3 3	0	0
3	A	1	Total C O 6 3 3	0	0
3	C	1	Total C O 6 3 3	0	0
3	C	1	Total C O 6 3 3	0	0
3	C	1	Total C O 6 3 3	0	0
3	C	1	Total C O 6 3 3	0	0
3	C	1	Total C O 6 3 3	0	0
3	D	1	Total C O 6 3 3	0	0

- Molecule 4 is POTASSIUM ION (three-letter code: K) (formula: K).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	3	Total K 3 3	0	0
4	C	3	Total K 3 3	0	0

- Molecule 5 is water.

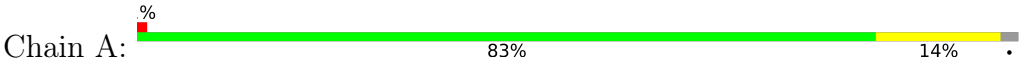
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	880	Total O 880 880	0	0
5	B	332	Total O 332 332	0	0
5	C	906	Total O 906 906	0	0
5	D	349	Total O 349 349	0	0

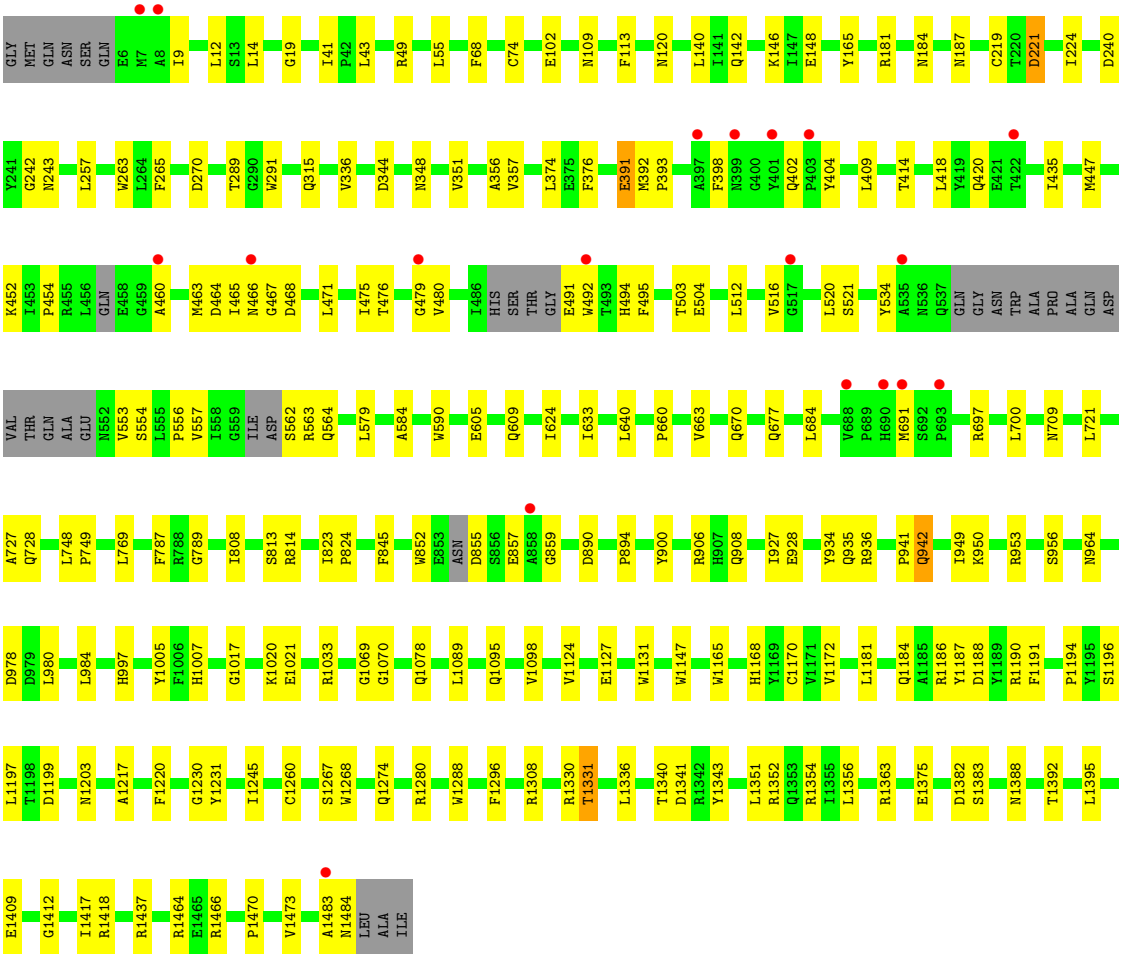
3 Residue-property plots (i)

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: YenB

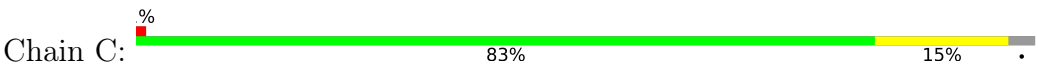
Chain A:

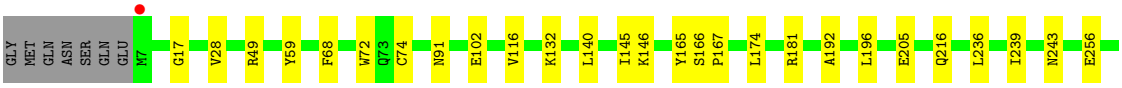


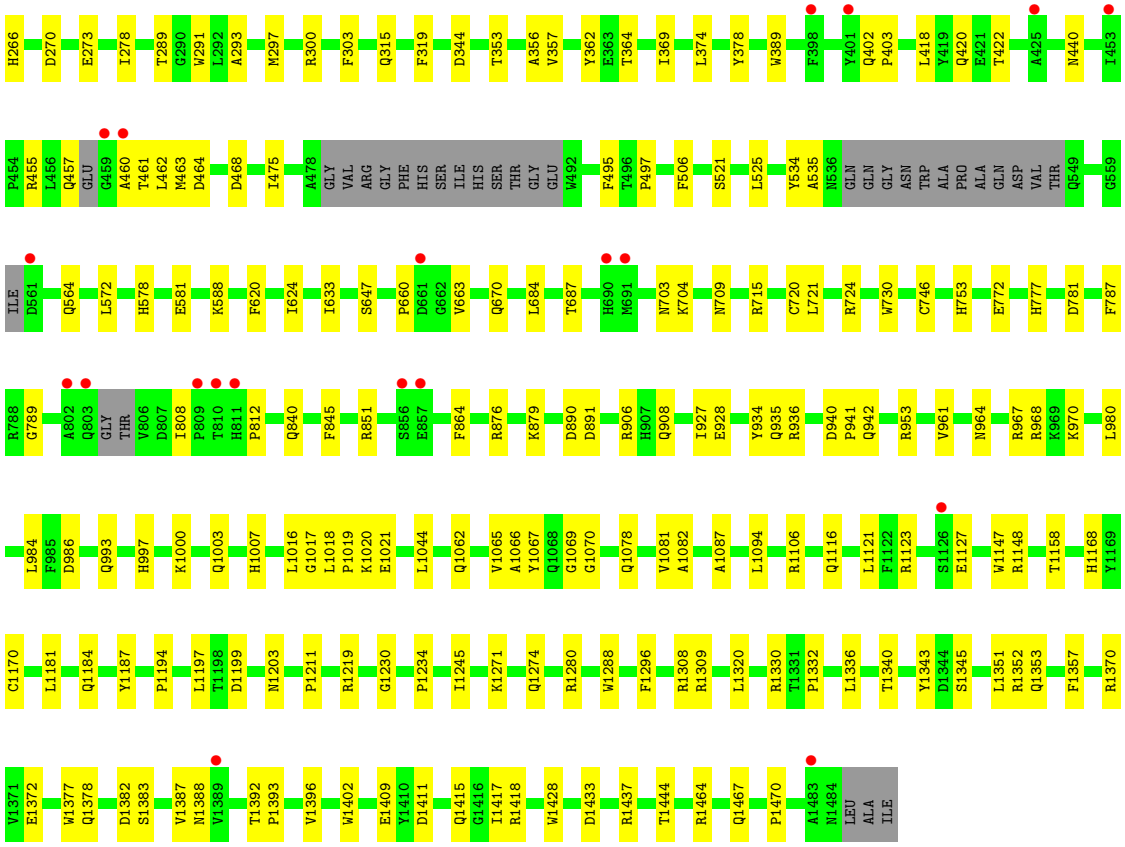


• Molecule 1: YenB

Chain C:

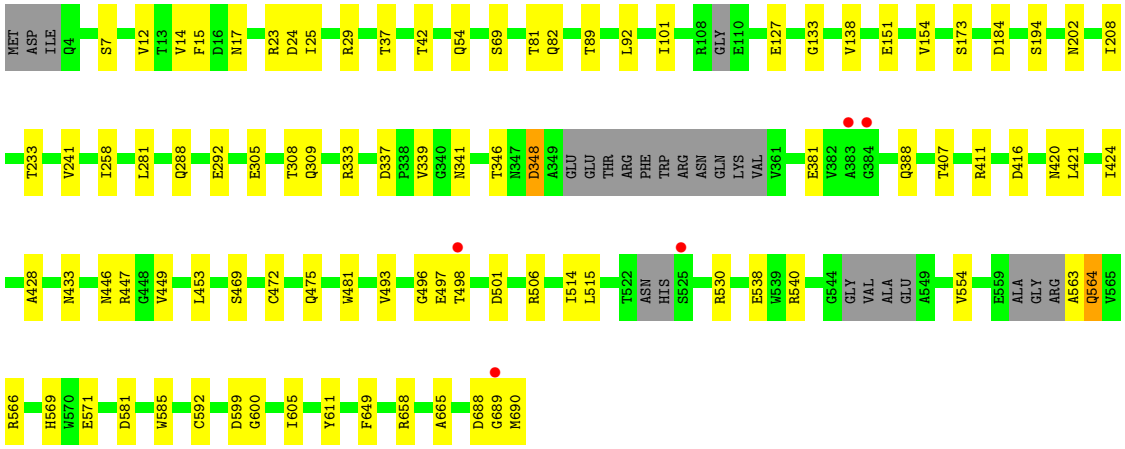






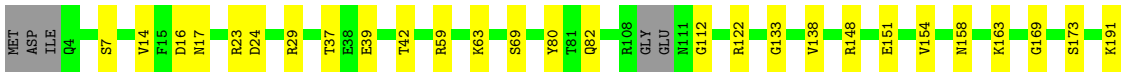
- Molecule 2: YenC2

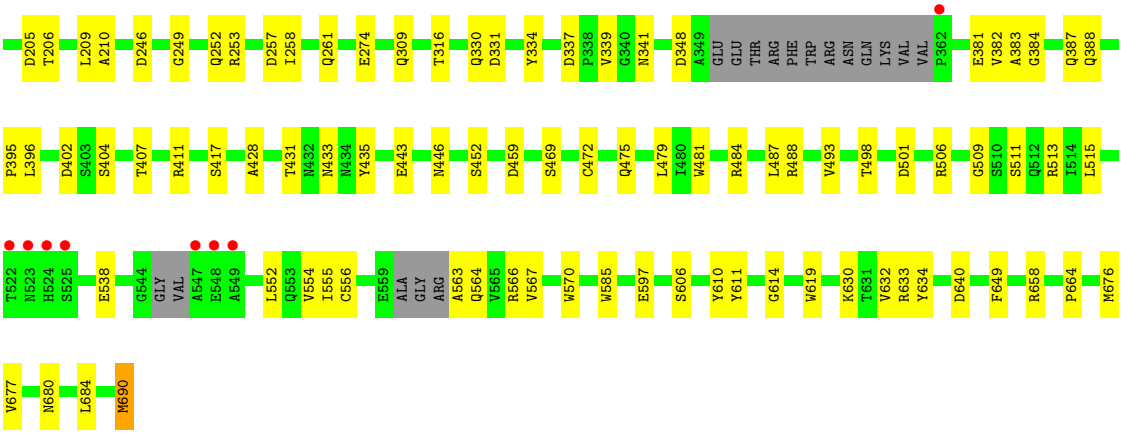
Chain B:



- Molecule 2: YenC2

Chain D:




D205 T206 L209 A210 D246 G249 Q252 R253 D257 I258 Q261 E274 Q309 T316 Q330 D331 Y334 D337 P338 V339 G340 N341 D348 A349 GLU GLU THR ARG PHE TRP ARG ASN GLN LYS VAL VAL P362 E381 V382 A383 G384 Q387 Q388
P395 L396 D402 S403 S404 T407 R411 S417 A428 T431 N432 N433 N434 Y435 E443 N446 S452 D459 S469 C472 Q475 L479 I480 W481 R484 L487 R488 V493 T498 D501 R506 G509 S510 S511 Q512 R513 I514 L515
T522 N523 H524 S525 E538 G544 GLY VAL A547 E548 A549 L552 Q553 V554 I555 C556 E559 ALA GLY ARG A563 Q564 V565 R566 V567 W570 W585 E597 S606 Y610 Y611 G614 W619 K630 T631 V632 R633 Y634 D640 F649 R658 P664 M676
V677 N680 L684 M690

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	133.72Å 147.60Å 274.35Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.55 – 2.49 49.55 – 2.26	Depositor EDS
% Data completeness (in resolution range)	100.0 (49.55-2.49) 88.9 (49.55-2.26)	Depositor EDS
R_{merge}	0.30	Depositor
R_{sym}	0.30	Depositor
$< I/\sigma(I) >$ [1]	0.24 (at 2.27Å)	Xtriage
Refinement program	PHENIX 1.8.1_1168	Depositor
R, R_{free}	0.200 , 0.251 0.202 , 0.250	Depositor DCC
R_{free} test set	12329 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	29.2	Xtriage
Anisotropy	0.093	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 41.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.43$, $< L^2 > = 0.25$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	35548	wwPDB-VP
Average B, all atoms (Å^2)	48.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.99% 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

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, K

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	0.29	0/11744	0.50	0/16018
1	C	0.29	0/11683	0.49	0/15938
2	B	0.27	0/5196	0.48	0/7081
2	D	0.27	0/5226	0.48	0/7119
All	All	0.29	0/33849	0.49	0/46156

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	11431	0	10793	128	0
1	C	11370	0	10711	143	0
2	B	5100	0	4844	50	0
2	D	5126	0	4894	75	0
3	A	12	0	16	1	0
3	C	30	0	40	2	0
3	D	6	0	8	1	0
4	A	3	0	0	0	0
4	C	3	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	A	880	0	0	13	0
5	B	332	0	0	6	0
5	C	906	0	0	29	0
5	D	349	0	0	12	0
All	All	35548	0	31306	385	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 385 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:A:49:ARG:NH2	1:A:243:ASN:O	2.15	0.80
2:D:14:VAL:HB	2:D:23:ARG:HB2	1.65	0.79
2:B:14:VAL:HB	2:B:23:ARG:HB2	1.62	0.79
2:D:472:CYS:SG	5:D:995:HOH:O	2.42	0.76
1:C:1271:LYS:NZ	5:C:1946:HOH:O	2.19	0.74

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	1448/1488 (97%)	1395 (96%)	52 (4%)	1 (0%)	51 73
1	C	1441/1488 (97%)	1387 (96%)	54 (4%)	0	100 100
2	B	654/690 (95%)	628 (96%)	24 (4%)	2 (0%)	41 61
2	D	659/690 (96%)	627 (95%)	30 (5%)	2 (0%)	41 61
All	All	4202/4356 (96%)	4037 (96%)	160 (4%)	5 (0%)	51 73

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	564	GLN
2	D	498	THR
1	A	468	ASP
2	D	112	GLY
2	B	496	GLY

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	1198/1266 (95%)	1184 (99%)	14 (1%)	71 88
1	C	1194/1266 (94%)	1191 (100%)	3 (0%)	92 97
2	B	534/579 (92%)	528 (99%)	6 (1%)	73 89
2	D	536/579 (93%)	533 (99%)	3 (1%)	86 95
All	All	3462/3690 (94%)	3436 (99%)	26 (1%)	81 93

5 of 26 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	339	VAL
2	B	447	ARG
2	D	585	TRP
2	B	407	THR
2	B	514	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	753	HIS
1	C	1259	HIS
2	D	261	GLN
2	B	288	GLN
1	A	262	GLN

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 14 ligands modelled in this entry, 6 are monoatomic - leaving 8 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	$\#\|Z\| > 2$
3	GOL	C	1505	-	5,5,5	0.33	0	5,5,5	0.30	0
3	GOL	D	701	-	5,5,5	0.36	0	5,5,5	0.16	0
3	GOL	C	1502	-	5,5,5	0.36	0	5,5,5	0.36	0
3	GOL	C	1501	-	5,5,5	0.37	0	5,5,5	0.09	0
3	GOL	C	1503	-	5,5,5	0.35	0	5,5,5	0.28	0
3	GOL	A	1502	-	5,5,5	0.34	0	5,5,5	0.41	0
3	GOL	A	1501	-	5,5,5	0.34	0	5,5,5	0.24	0
3	GOL	C	1504	-	5,5,5	0.32	0	5,5,5	0.76	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
3	GOL	C	1505	-	-	1/4/4/4	-
3	GOL	D	701	-	-	1/4/4/4	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	GOL	C	1502	-	-	2/4/4/4	-
3	GOL	C	1501	-	-	4/4/4/4	-
3	GOL	C	1503	-	-	2/4/4/4	-
3	GOL	A	1502	-	-	1/4/4/4	-
3	GOL	A	1501	-	-	3/4/4/4	-
3	GOL	C	1504	-	-	0/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 14 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	1501	GOL	O1-C1-C2-C3
3	C	1501	GOL	O1-C1-C2-O2
3	C	1501	GOL	O1-C1-C2-C3
3	C	1502	GOL	O1-C1-C2-O2
3	C	1502	GOL	O1-C1-C2-C3

There are no ring outliers.

4 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	D	701	GOL	1	0
3	C	1503	GOL	1	0
3	A	1502	GOL	1	0
3	C	1504	GOL	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	A	1457/1488 (97%)	-0.20	19 (1%)	77	79	35, 46, 69, 87	0
1	C	1449/1488 (97%)	-0.18	21 (1%)	75	77	35, 45, 66, 84	0
2	B	666/690 (96%)	-0.19	5 (0%)	86	87	40, 50, 66, 89	0
2	D	668/690 (96%)	-0.22	8 (1%)	79	80	40, 50, 67, 95	0
All	All	4240/4356 (97%)	-0.20	53 (1%)	77	79	35, 47, 67, 95	0

The worst 5 of 53 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	856	SER	7.7
1	C	7	MET	5.1
1	A	690	HIS	4.6
2	B	383	ALA	4.5
1	A	397	ALA	4.1

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

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
3	GOL	C	1505	6/6	0.76	0.16	50,52,55,56	0
3	GOL	C	1502	6/6	0.83	0.18	47,48,51,53	0
3	GOL	C	1503	6/6	0.84	0.25	59,61,62,62	0
3	GOL	C	1504	6/6	0.86	0.18	43,46,48,51	0
3	GOL	A	1502	6/6	0.88	0.24	46,47,49,49	0
3	GOL	C	1501	6/6	0.88	0.12	55,56,58,60	0
3	GOL	A	1501	6/6	0.88	0.14	49,50,51,52	0
3	GOL	D	701	6/6	0.89	0.13	48,50,52,52	0
4	K	A	1505	1/1	0.90	0.18	57,57,57,57	0
4	K	C	1508	1/1	0.93	0.13	56,56,56,56	0
4	K	A	1503	1/1	0.97	0.13	45,45,45,45	0
4	K	C	1507	1/1	0.98	0.05	50,50,50,50	0
4	K	A	1504	1/1	0.98	0.05	49,49,49,49	0
4	K	C	1506	1/1	0.99	0.18	39,39,39,39	0

6.5 Other polymers (i)

There are no such residues in this entry.