wwPDB X-ray Structure Validation Summary Report (i)

Oct 11, 2023 – 04:31 PM EDT

PDB ID	:	7RYO
Title	:	CD1a-dideoxymycobactin-gdTCR complex
Authors	:	Wegrecki, M.; Le Nours, J.; Rossjohn, J.
Deposited on	:	2021-08-25
Resolution	:	3.00 Å(reported)

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	:	2.35.1
buster-report	:	1.1.7 (2018)
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)	:	2.35.1

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 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.
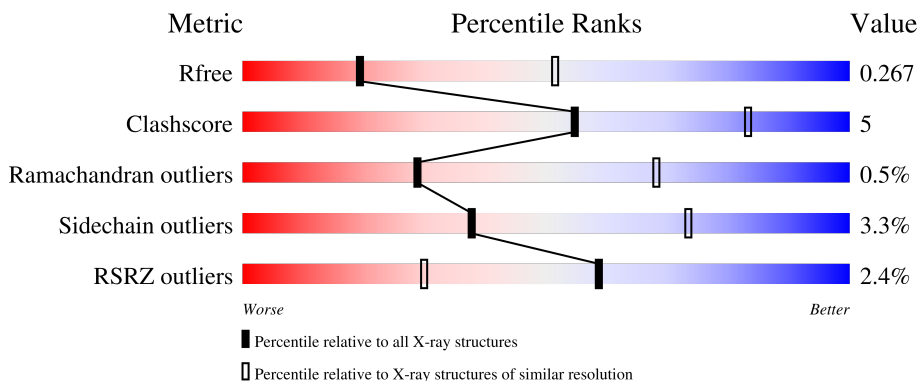
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)

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 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	286	85% · · · 12% ·
2	B	108	75% · · · 14% · 8%
3	C	248	2% · 77% · · · 18% · ·
4	D	209	7% · 78% · · · 13% · 8%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard

residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
5	EDO	A	402	-	-	-	X

2 Entry composition ⓘ

There are 7 unique types of molecules in this entry. The entry contains 6288 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 T-cell surface glycoprotein CD1a.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	277	2195	1410	382	395	8	0	0	0

There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	ASP	-	expression tag	UNP P06126
A	0	ALA	-	expression tag	UNP P06126
A	2	THR	ASP	conflict	UNP P06126
A	13	ILE	THR	variant	UNP P06126
A	51	TRP	CYS	variant	UNP P06126
A	279	GLY	-	expression tag	UNP P06126
A	280	SER	-	expression tag	UNP P06126
A	281	LEU	-	expression tag	UNP P06126
A	282	VAL	-	expression tag	UNP P06126
A	283	PRO	-	expression tag	UNP P06126
A	284	ARG	-	expression tag	UNP P06126

- Molecule 2 is a protein called Beta-2-microglobulin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	99	812	516	137	157	2	0	0	0

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-1	ASP	-	expression tag	UNP P61769
B	0	ALA	-	expression tag	UNP P61769
B	1	GLY	-	expression tag	UNP P61769
B	101	GLY	-	expression tag	UNP P61769
B	102	SER	-	expression tag	UNP P61769

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	103	LEU	-	expression tag	UNP P61769
B	104	VAL	-	expression tag	UNP P61769
B	105	PRO	-	expression tag	UNP P61769
B	106	ARG	-	expression tag	UNP P61769

- Molecule 3 is a protein called T cell receptor gamma variable 4,T cell receptor beta constant 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	238	Total 1817	C 1155	N 311	O 345	S 6	0	0	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	1	MET	-	initiating methionine	UNP A0A0C4DH28
C	2	ALA	-	expression tag	UNP A0A0C4DH28
C	103	ASP	-	linker	UNP A0A0C4DH28
C	104	TYR	-	linker	UNP A0A0C4DH28
C	105	TYR	-	linker	UNP A0A0C4DH28
C	106	LYS	-	linker	UNP A0A0C4DH28
C	107	LYS	-	linker	UNP A0A0C4DH28
C	108	LEU	-	linker	UNP A0A0C4DH28
C	109	PHE	-	linker	UNP A0A0C4DH28
C	110	GLY	-	linker	UNP A0A0C4DH28
C	111	SER	-	linker	UNP A0A0C4DH28
C	112	GLY	-	linker	UNP A0A0C4DH28
C	113	THR	-	linker	UNP A0A0C4DH28
C	114	THR	-	linker	UNP A0A0C4DH28
C	115	LEU	-	linker	UNP A0A0C4DH28
C	116	VAL	-	linker	UNP A0A0C4DH28
C	117	VAL	-	linker	UNP A0A0C4DH28
C	118	THR	-	linker	UNP A0A0C4DH28
C	119	GLU	-	linker	UNP A0A0C4DH28
C	122	LYS	ASN	conflict	UNP P01850
C	123	ASN	LYS	conflict	UNP P01850
C	155	TYR	PHE	conflict	UNP P01850
C	175	CYS	SER	engineered mutation	UNP P01850
C	193	ALA	CYS	engineered mutation	UNP P01850

- Molecule 4 is a protein called T cell receptor delta variable 1,T cell receptor alpha chain constant.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	D	192	1391	873	233	278	7	0	0	0

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	1	MET	-	initiating methionine	UNP A0A1B0GX56
D	97	LEU	-	linker	UNP A0A1B0GX56
D	98	ARG	-	linker	UNP A0A1B0GX56
D	99	TRP	-	linker	UNP A0A1B0GX56
D	100	PRO	-	linker	UNP A0A1B0GX56
D	101	ASP	-	linker	UNP A0A1B0GX56
D	102	LYS	-	linker	UNP A0A1B0GX56
D	103	LEU	-	linker	UNP A0A1B0GX56
D	104	ILE	-	linker	UNP A0A1B0GX56
D	105	PHE	-	linker	UNP A0A1B0GX56
D	106	GLY	-	linker	UNP A0A1B0GX56
D	107	LYS	-	linker	UNP A0A1B0GX56
D	108	GLY	-	linker	UNP A0A1B0GX56
D	109	THR	-	linker	UNP A0A1B0GX56
D	110	ARG	-	linker	UNP A0A1B0GX56
D	111	VAL	-	linker	UNP A0A1B0GX56
D	112	THR	-	linker	UNP A0A1B0GX56
D	113	VAL	-	linker	UNP A0A1B0GX56
D	114	GLU	-	linker	UNP A0A1B0GX56
D	115	PRO	-	linker	UNP A0A1B0GX56
D	116	ASN	-	linker	UNP A0A1B0GX56
D	163	CYS	THR	engineered mutation	UNP P01848

- Molecule 5 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).

EDO

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total 4	C 2	O 2	0	0
5	A	1	Total 4	C 2	O 2	0	0

- Molecule 6 is dideoxymycobactin-838 (three-letter code: WGR) (formula: $C_{47}H_{75}N_5O_8$) (labeled as "Ligand of Interest" by depositor).

WGR
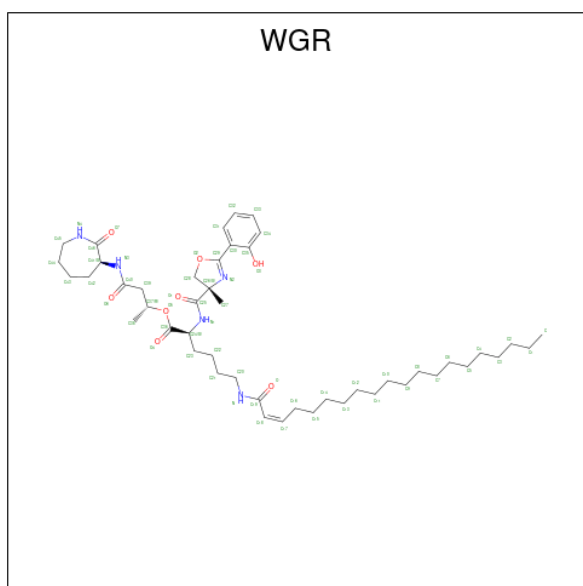
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
6	A	1	Total 60	C 47	N 5	O 8	0	0

• Molecule 7 is SULFATE ION (three-letter code: SO4) (formula: O_4S).
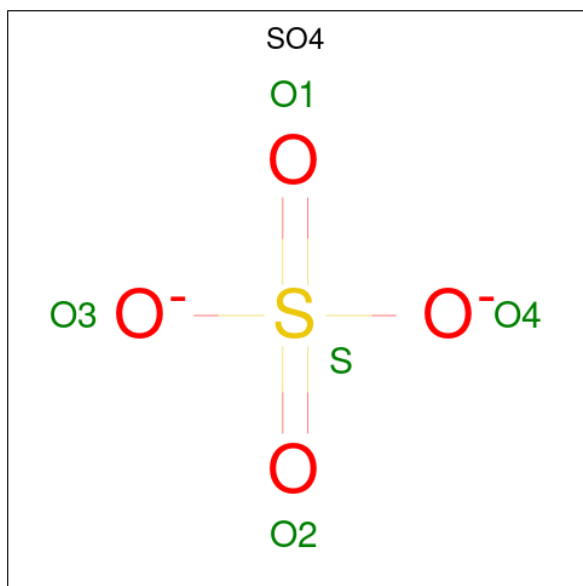
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
7	D	1	Total 5	O 4	S 1	0	0

# 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.
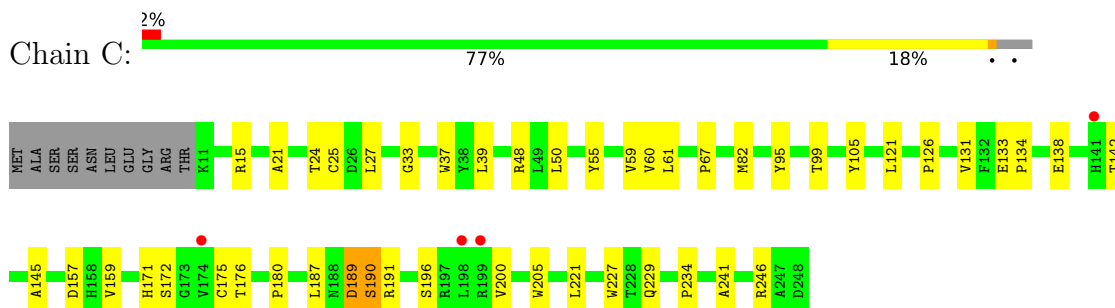
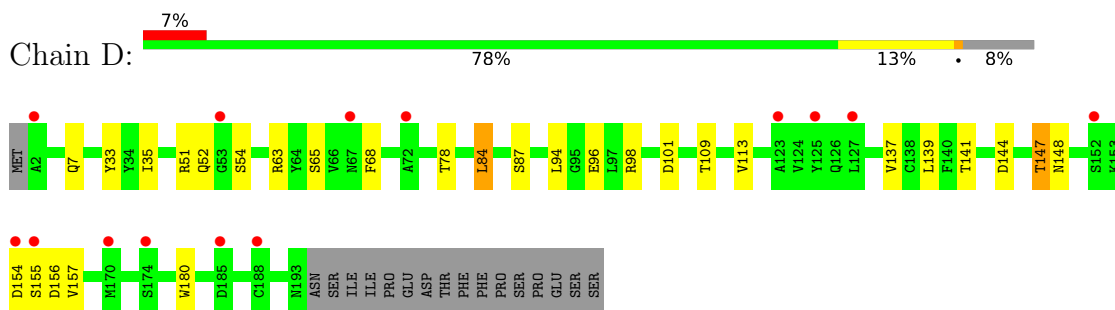• Molecule 1: T-cell surface glycoprotein CD1a
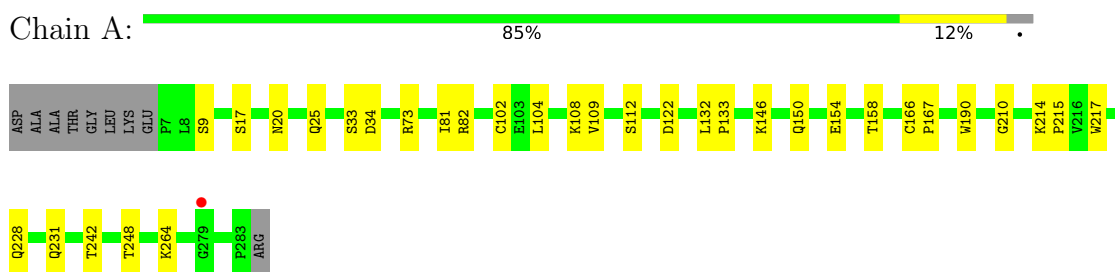
Chain A:

• Molecule 2: Beta-2-microglobulin
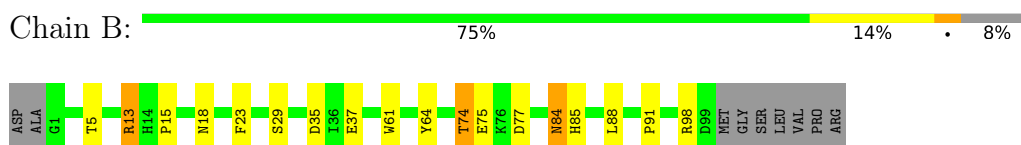
Chain B:

• Molecule 3: T cell receptor gamma variable 4,T cell receptor beta constant 1

Chain C:

• Molecule 4: T cell receptor delta variable 1,T cell receptor alpha chain constant

Chain D:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	234.88Å 42.24Å 124.66Å 90.00° 113.35° 90.00°	Depositor
Resolution (Å)	39.96 – 3.00 41.51 – 3.00	Depositor EDS
% Data completeness (in resolution range)	99.4 (39.96-3.00) 99.5 (41.51-3.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.87 (at 3.01Å)	Xtriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, R_{free}	0.222 , 0.267 0.221 , 0.267	Depositor DCC
R_{free} test set	1153 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	61.2	Xtriage
Anisotropy	0.754	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 47.5	EDS
L-test for twinning[2]	$<|L|> = 0.48$, $<L^2> = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	6288	wwPDB-VP
Average B, all atoms (Å^2)	79.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.95% 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: SO4, EDO, WGR

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	0.24	0/2268	0.46	0/3094
2	B	0.24	0/835	0.47	0/1135
3	C	0.24	0/1868	0.49	0/2559
4	D	0.24	0/1419	0.49	0/1939
All	All	0.24	0/6390	0.48	0/8727

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	2195	0	2047	19	0
2	B	812	0	755	9	0
3	C	1817	0	1675	25	0
4	D	1391	0	1238	16	0
5	A	8	0	12	1	0
6	A	60	0	0	1	0
7	D	5	0	0	0	0
All	All	6288	0	5727	62	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 5.

The worst 5 of 62 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:217:TRP:HB3	1:A:264:LYS:HB2	1.73	0.69
4:D:84:LEU:HA	4:D:113:VAL:HG21	1.76	0.68
3:C:48:ARG:NH1	4:D:101:ASP:OD1	2.33	0.61
4:D:144:ASP:OD1	4:D:147:THR:OG1	2.18	0.61
1:A:9:SER:O	1:A:33:SER:OG	2.19	0.60

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	275/286 (96%)	256 (93%)	18 (6%)	1 (0%)	34	72
2	B	97/108 (90%)	93 (96%)	4 (4%)	0	100	100
3	C	236/248 (95%)	225 (95%)	11 (5%)	0	100	100
4	D	190/209 (91%)	168 (88%)	19 (10%)	3 (2%)	9	40
All	All	798/851 (94%)	742 (93%)	52 (6%)	4 (0%)	29	68

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	109	VAL
4	D	155	SER
4	D	154	ASP
4	D	157	VAL

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	229/247 (93%)	228 (100%)	1 (0%)	91	97
2	B	90/100 (90%)	84 (93%)	6 (7%)	16	49
3	C	187/215 (87%)	179 (96%)	8 (4%)	29	66
4	D	136/188 (72%)	130 (96%)	6 (4%)	28	65
All	All	642/750 (86%)	621 (97%)	21 (3%)	38	73

5 of 21 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	C	191	ARG
4	D	84	LEU
4	D	156	ASP
4	D	147	THR
4	D	68	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
4	D	7	GLN

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 (i)

4 ligands are 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	$\#|Z| > 2$
5	EDO	A	401	-	3,3,3	0.46	0	2,2,2	0.34	0
5	EDO	A	402	-	3,3,3	0.46	0	2,2,2	0.33	0
6	WGR	A	403	-	61,62,62	0.15	0	71,78,78	0.36	0
7	SO4	D	301	-	4,4,4	0.14	0	6,6,6	0.18	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
5	EDO	A	401	-	-	0/1/1/1	-
5	EDO	A	402	-	-	0/1/1/1	-
6	WGR	A	403	-	-	18/57/80/80	0/3/3/3

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 18 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
6	A	403	WGR	O-C19-N-C20
6	A	403	WGR	C18-C19-N-C20
6	A	403	WGR	C22-C23-C24-N1
6	A	403	WGR	C39-C40-N3-C41
6	A	403	WGR	O6-C40-N3-C41

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	402	EDO	1	0
6	A	403	WGR	1	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand WGR A 403

Bond lengths

Bond angles

Torsions

Rings
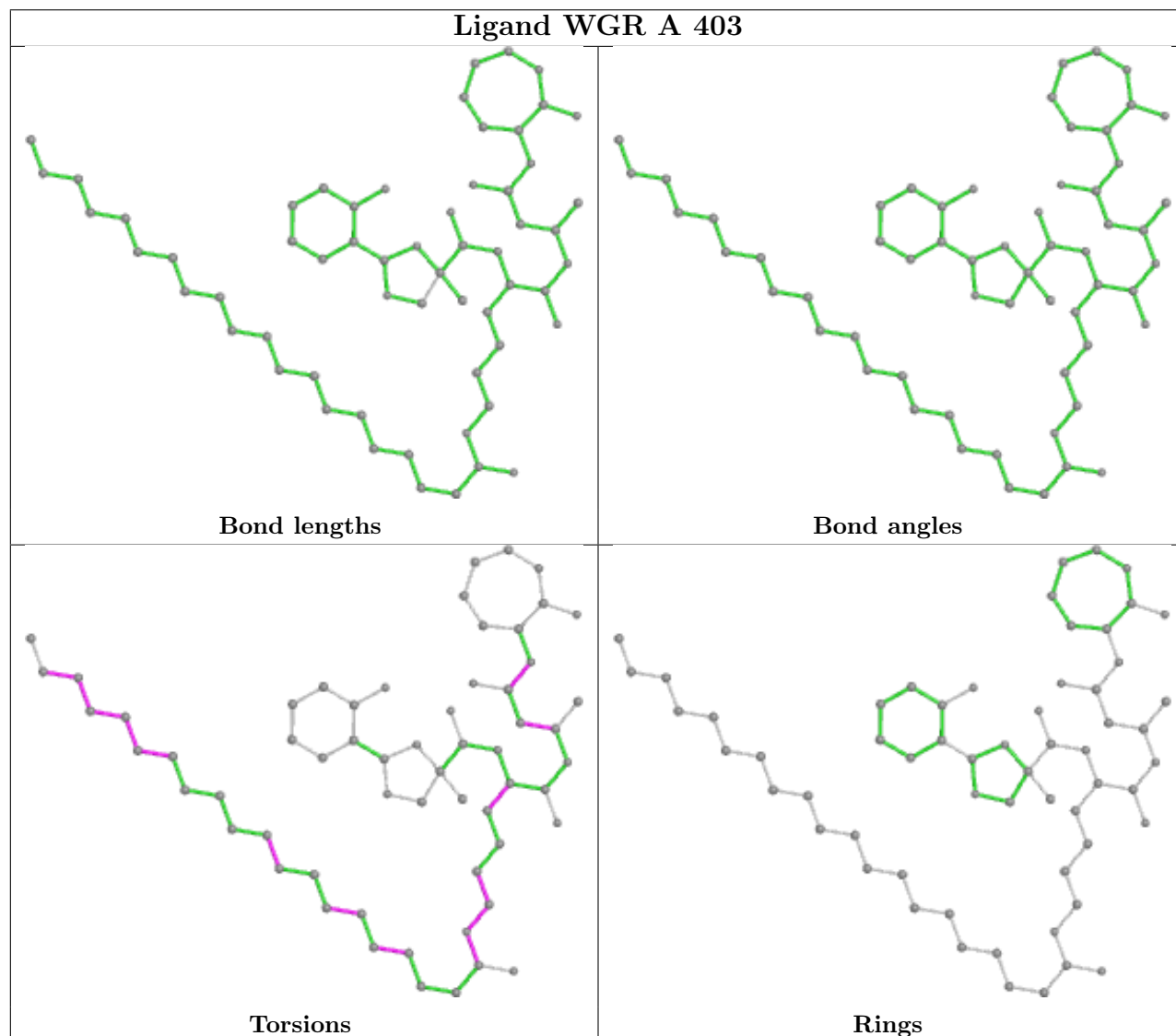
5.7 Other polymers ⓘ

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(Å^2)	Q<0.9
1	A	277/286 (96%)	-0.20	1 (0%)	92	79	39, 61, 102, 135	0
2	B	99/108 (91%)	-0.39	0	100	100	34, 51, 83, 106	0
3	C	238/248 (95%)	-0.13	4 (1%)	70	41	51, 89, 125, 149	0
4	D	192/209 (91%)	0.29	14 (7%)	15	4	56, 112, 155, 167	0
All	All	806/851 (94%)	-0.09	19 (2%)	59	30	34, 76, 133, 167	0

The worst 5 of 19 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	199	ARG	3.6
3	C	174	VAL	3.3
1	A	279	GLY	3.1
4	D	2	ALA	3.1
4	D	152	SER	3.0

6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

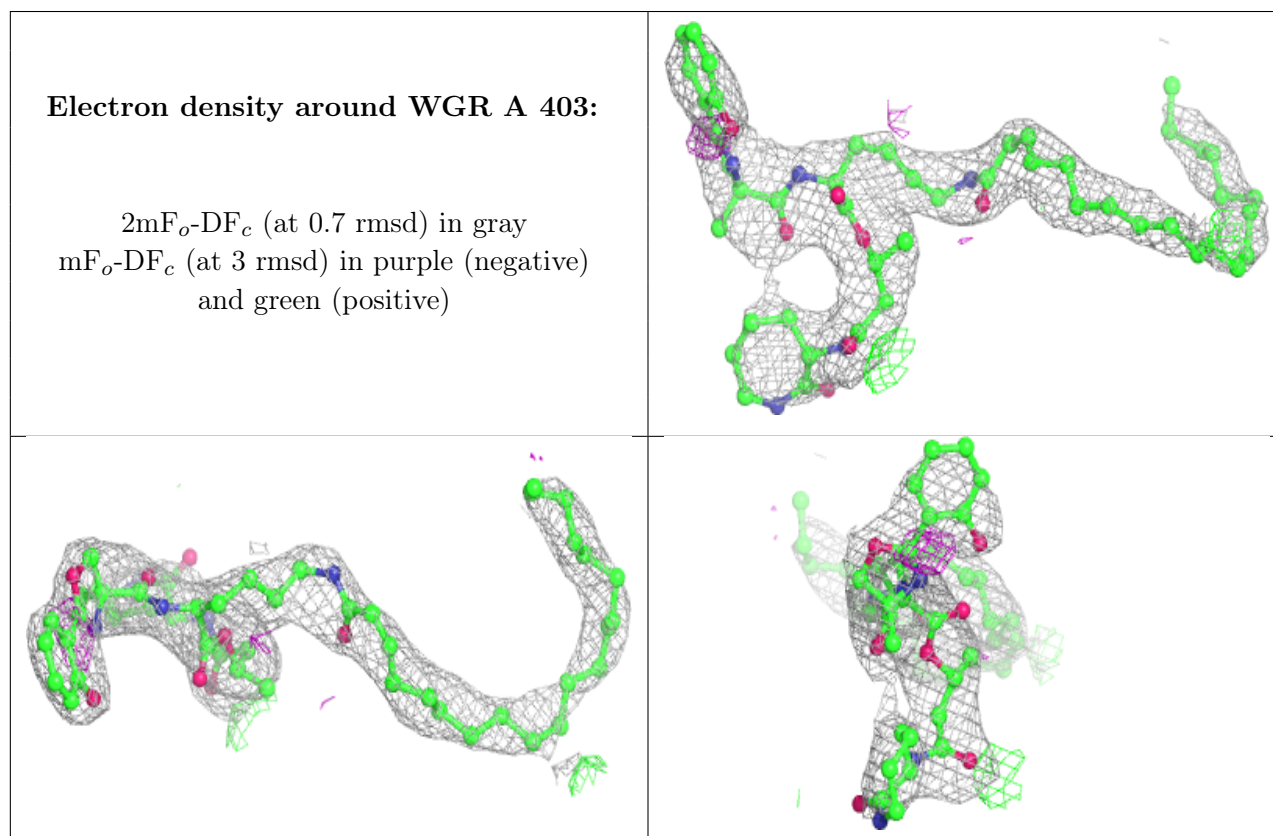There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

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, 95th 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
5	EDO	A	402	4/4	0.75	0.58	53,55,60,67	0
7	SO4	D	301	5/5	0.87	0.18	83,89,110,159	0
5	EDO	A	401	4/4	0.88	0.32	40,40,42,44	0
6	WGR	A	403	60/60	0.91	0.39	30,66,92,108	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around WGR A 403:

2mF$_o$-DF$_c$ (at 0.7 rmsd) in gray
mF$_o$-DF$_c$ (at 3 rmsd) in purple (negative)
and green (positive)

6.5 Other polymers ⓘ

There are no such residues in this entry.