



wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 11:38 am BST

PDB ID : 6U8J
Title : Crystal structure of 3-deoxy-D-arabinoheptulosonate-7-phosphate synthase/p
hospho-2-dehydro-3-deoxyheptonate aldolase (Aro3) from *Candida auris*
Authors : Michalska, K.; Evdokimova, E.; Semper, C.; Di Leo, R.; Stogios, P.J.;
Savchenko, A.; Joachimiak, A.; Center for Structural Genomics of Infectious
Diseases (CSGID)
Deposited on : 2019-09-05
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 ⓘ symbol.

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

MolProbity	:	4.02b-467
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

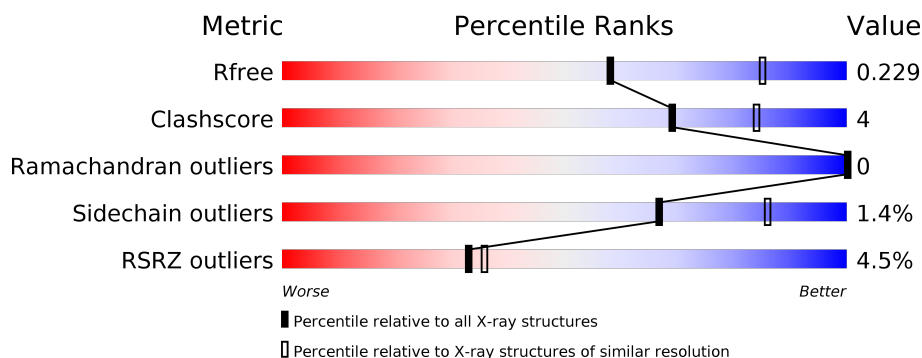
1 Overall quality at a glance

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 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 $\leq 5\%$. 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	376	<div> <div>3%</div> <div>90%</div> <div>9%</div> <div>.</div> </div>
1	B	376	<div> <div>%</div> <div>87%</div> <div>12%</div> <div>.</div> </div>
1	C	376	<div> <div>86%</div> <div>11%</div> <div>..</div> </div>
1	D	376	<div> <div>%</div> <div>93%</div> <div>6%</div> <div>.</div> </div>
1	E	376	<div> <div>8%</div> <div>86%</div> <div>12%</div> <div>..</div> </div>
1	F	376	<div> <div>9%</div> <div>87%</div> <div>11%</div> <div>..</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	376	<div><div></div><div>7%</div><div></div><div>88%</div><div></div><div>10%</div><div></div></div>
1	H	376	<div><div></div><div>6%</div><div></div><div>84%</div><div></div><div>11%</div><div></div><div>5%</div></div>

2 Entry composition i

There are 2 unique types of molecules in this entry. The entry contains 22717 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 Phospho-2-dehydro-3-deoxyheptonate aldolase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	372	Total	C	N	O	S	0	0	0
			2855	1780	512	549	14			
1	B	372	Total	C	N	O	S	0	0	0
			2855	1780	512	549	14			
1	C	369	Total	C	N	O	S	0	0	0
			2839	1771	510	544	14			
1	D	374	Total	C	N	O	S	0	0	0
			2870	1789	515	552	14			
1	E	372	Total	C	N	O	S	0	0	0
			2855	1780	512	549	14			
1	F	371	Total	C	N	O	S	0	0	0
			2847	1774	511	548	14			
1	G	369	Total	C	N	O	S	0	0	0
			2830	1766	506	544	14			
1	H	358	Total	C	N	O	S	0	0	0
			2762	1721	496	531	14			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	SER	-	expression tag	UNP A0A2H0ZWN3
A	-1	ASN	-	expression tag	UNP A0A2H0ZWN3
A	0	ALA	-	expression tag	UNP A0A2H0ZWN3
B	-2	SER	-	expression tag	UNP A0A2H0ZWN3
B	-1	ASN	-	expression tag	UNP A0A2H0ZWN3
B	0	ALA	-	expression tag	UNP A0A2H0ZWN3
C	-2	SER	-	expression tag	UNP A0A2H0ZWN3
C	-1	ASN	-	expression tag	UNP A0A2H0ZWN3
C	0	ALA	-	expression tag	UNP A0A2H0ZWN3
D	-2	SER	-	expression tag	UNP A0A2H0ZWN3
D	-1	ASN	-	expression tag	UNP A0A2H0ZWN3
D	0	ALA	-	expression tag	UNP A0A2H0ZWN3
E	-2	SER	-	expression tag	UNP A0A2H0ZWN3

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	-1	ASN	-	expression tag	UNP A0A2H0ZWN3
E	0	ALA	-	expression tag	UNP A0A2H0ZWN3
F	-2	SER	-	expression tag	UNP A0A2H0ZWN3
F	-1	ASN	-	expression tag	UNP A0A2H0ZWN3
F	0	ALA	-	expression tag	UNP A0A2H0ZWN3
G	-2	SER	-	expression tag	UNP A0A2H0ZWN3
G	-1	ASN	-	expression tag	UNP A0A2H0ZWN3
G	0	ALA	-	expression tag	UNP A0A2H0ZWN3
H	-2	SER	-	expression tag	UNP A0A2H0ZWN3
H	-1	ASN	-	expression tag	UNP A0A2H0ZWN3
H	0	ALA	-	expression tag	UNP A0A2H0ZWN3

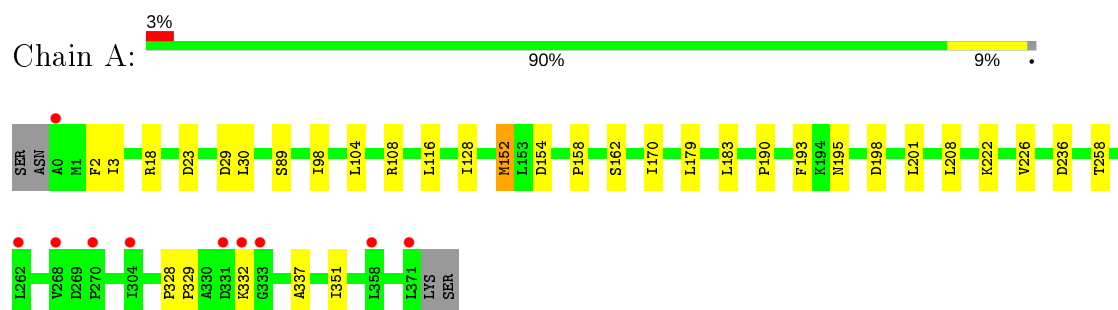
- Molecule 2 is UNKNOWN ATOM OR ION (three-letter code: UNX) (formula: X).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	1	Total X 1 1	0	0
2	A	1	Total X 1 1	0	0
2	D	1	Total X 1 1	0	0
2	C	1	Total X 1 1	0	0

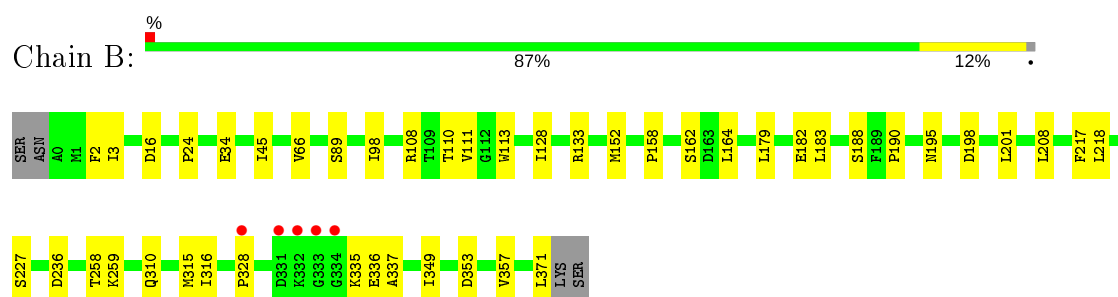
3 Residue-property plots [i](#)

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.

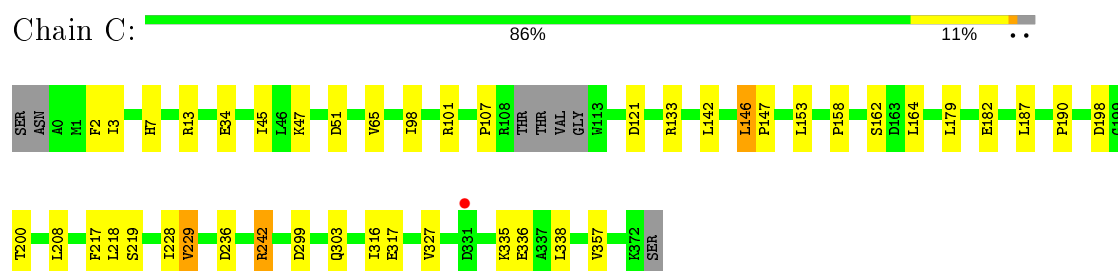
- Molecule 1: Phospho-2-dehydro-3-deoxyheptonate aldolase



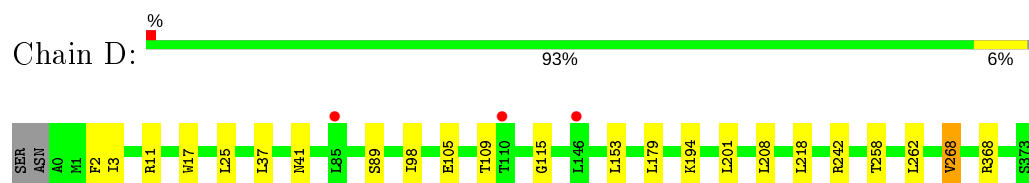
- Molecule 1: Phospho-2-dehydro-3-deoxyheptonate aldolase



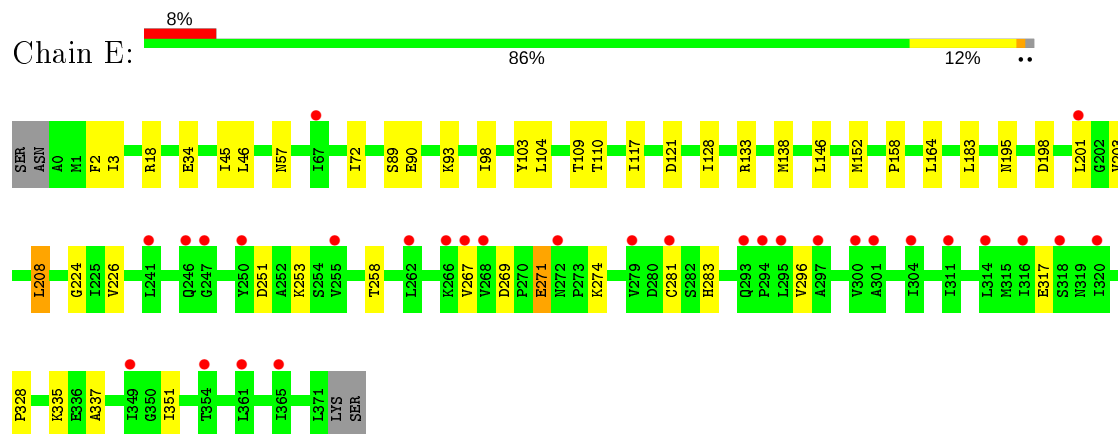
- Molecule 1: Phospho-2-dehydro-3-deoxyheptonate aldolase



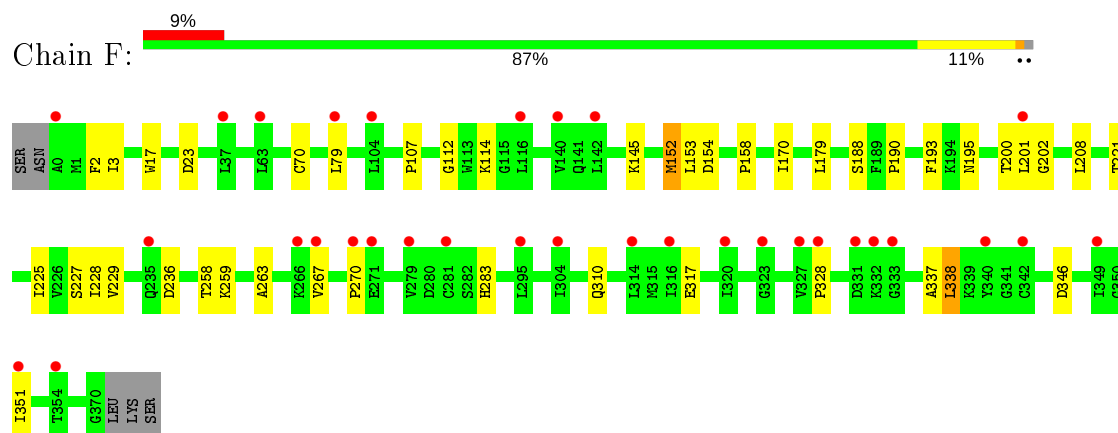
- Molecule 1: Phospho-2-dehydro-3-deoxyheptonate aldolase



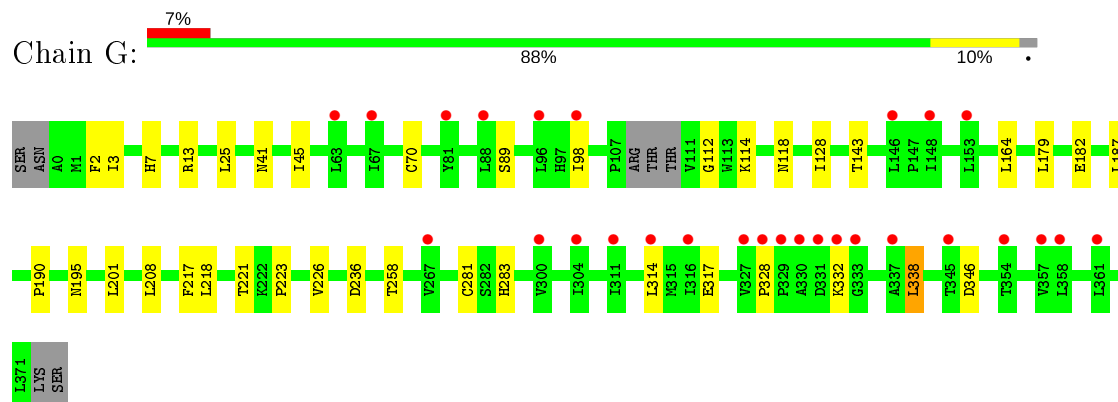
- Molecule 1: Phospho-2-dehydro-3-deoxyheptonate aldolase



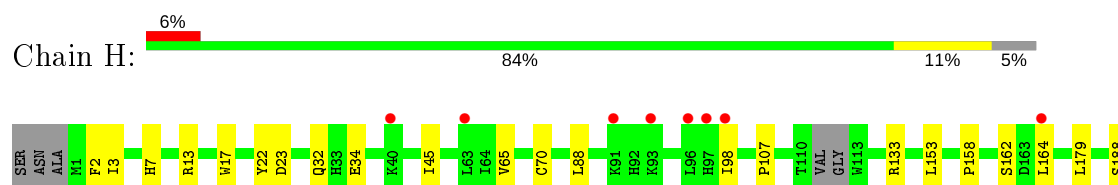
- Molecule 1: Phospho-2-dehydro-3-deoxyheptonate aldolase

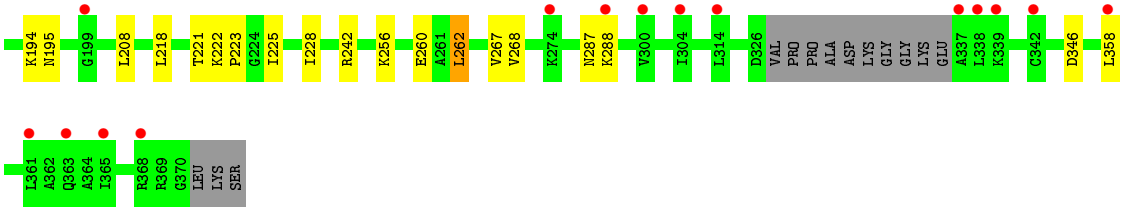


- Molecule 1: Phospho-2-dehydro-3-deoxyheptonate aldolase



- Molecule 1: Phospho-2-dehydro-3-deoxyheptonate aldolase





4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	62.50Å 141.35Å 192.87Å 90.00° 97.92° 90.00°	Depositor
Resolution (Å)	48.97 – 2.49 48.96 – 2.49	Depositor EDS
% Data completeness (in resolution range)	97.8 (48.97-2.49) 97.9 (48.96-2.49)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.89 (at 2.48Å)	Xtriage
Refinement program	PHENIX dev_2947	Depositor
R, R_{free}	0.184 , 0.229 0.184 , 0.229	Depositor DCC
R_{free} test set	5683 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å ²)	64.1	Xtriage
Anisotropy	0.657	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 60.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.027 for h,-k,-h-l	Xtriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	22717	wwPDB-VP
Average B, all atoms (Å ²)	90.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.96% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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: UNX

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	0.43	0/2904	0.63	0/3920
1	B	0.45	0/2904	0.64	1/3920 (0.0%)
1	C	0.50	0/2887	0.67	1/3893 (0.0%)
1	D	0.42	0/2919	0.64	1/3939 (0.0%)
1	E	0.37	0/2904	0.57	0/3920
1	F	0.35	0/2896	0.57	0/3909
1	G	0.35	0/2878	0.57	0/3883
1	H	0.35	0/2807	0.57	1/3785 (0.0%)
All	All	0.41	0/23099	0.61	4/31169 (0.0%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	H	262	LEU	CA-CB-CG	5.42	127.77	115.30
1	D	368	ARG	NE-CZ-NH2	-5.30	117.65	120.30
1	C	146	LEU	CA-CB-CG	5.23	127.33	115.30
1	B	16	ASP	CB-CG-OD2	5.02	122.82	118.30

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	2855	0	2867	21	0
1	B	2855	0	2867	26	0
1	C	2839	0	2853	28	0
1	D	2870	0	2886	13	0
1	E	2855	0	2868	27	0
1	F	2847	0	2857	28	0
1	G	2830	0	2840	21	0
1	H	2762	0	2768	25	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
All	All	22717	0	22806	166	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 4.

The worst 5 of 166 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:201:LEU:HD21	1:A:258:THR:HG23	1.60	0.83
1:D:194:LYS:HD3	1:D:242:ARG:HD2	1.64	0.80
1:H:194:LYS:HD3	1:H:242:ARG:HD2	1.68	0.76
1:G:328:PRO:HD3	1:G:338:LEU:HD13	1.70	0.71
1:B:201:LEU:HD21	1:B:258:THR:HG23	1.75	0.69

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	370/376 (98%)	358 (97%)	12 (3%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	370/376 (98%)	359 (97%)	11 (3%)	0	100	100
1	C	365/376 (97%)	354 (97%)	11 (3%)	0	100	100
1	D	372/376 (99%)	358 (96%)	14 (4%)	0	100	100
1	E	370/376 (98%)	359 (97%)	11 (3%)	0	100	100
1	F	369/376 (98%)	357 (97%)	12 (3%)	0	100	100
1	G	365/376 (97%)	353 (97%)	12 (3%)	0	100	100
1	H	352/376 (94%)	342 (97%)	10 (3%)	0	100	100
All	All	2933/3008 (98%)	2840 (97%)	93 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	311/315 (99%)	304 (98%)	7 (2%)	50	76
1	B	311/315 (99%)	306 (98%)	5 (2%)	62	84
1	C	309/315 (98%)	305 (99%)	4 (1%)	69	87
1	D	313/315 (99%)	310 (99%)	3 (1%)	76	90
1	E	311/315 (99%)	306 (98%)	5 (2%)	62	84
1	F	310/315 (98%)	307 (99%)	3 (1%)	76	90
1	G	308/315 (98%)	305 (99%)	3 (1%)	76	90
1	H	302/315 (96%)	298 (99%)	4 (1%)	69	87
All	All	2475/2520 (98%)	2441 (99%)	34 (1%)	67	86

5 of 34 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	336	GLU
1	E	146	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	H	208	LEU
1	D	208	LEU
1	A	226	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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](#)

Of 4 ligands modelled in this entry, 4 are unknown - leaving 0 for Mogul analysis.

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 [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, 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	372/376 (98%)	0.15	10 (2%) 54 58	52, 75, 110, 161	0
1	B	372/376 (98%)	0.11	5 (1%) 77 79	52, 70, 104, 147	0
1	C	369/376 (98%)	0.08	1 (0%) 94 94	49, 67, 94, 146	0
1	D	374/376 (99%)	0.12	3 (0%) 86 87	58, 80, 108, 146	0
1	E	372/376 (98%)	0.48	30 (8%) 12 12	65, 94, 131, 172	0
1	F	371/376 (98%)	0.51	32 (8%) 10 10	75, 101, 138, 189	0
1	G	369/376 (98%)	0.38	28 (7%) 13 14	73, 104, 144, 181	0
1	H	358/376 (95%)	0.54	23 (6%) 19 20	72, 108, 143, 173	0
All	All	2957/3008 (98%)	0.30	132 (4%) 33 36	49, 87, 132, 189	0

The worst 5 of 132 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	327	VAL	6.3
1	H	361	LEU	5.9
1	G	333	GLY	5.9
1	E	361	LEU	5.8
1	E	316	ILE	5.2

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 [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, 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
2	UNX	A	401	1/1	0.67	0.14	44,44,44,44	0
2	UNX	D	401	1/1	0.87	0.21	58,58,58,58	0
2	UNX	C	401	1/1	0.90	0.18	48,48,48,48	0
2	UNX	B	401	1/1	0.93	0.15	50,50,50,50	0

6.5 Other polymers [i](#)

There are no such residues in this entry.