



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 9, 2018 – 08:29 am GMT

PDB ID : 2C8N
Title : The Structure of a family 51 arabinofuranosidase, Araf51, from *Clostridium thermocellum* in complex with 1,3-linked arabinoside of xylobiose.
Authors : Taylor, E.J.; Smith, N.L.; Turkenburg, J.P.; D'Souza, S.; Gilbert, H.J.; Davies, G.J.
Deposited on : 2005-12-06
Resolution : 2.90 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
Mogul : 1.7.3 (157068), CSD as539be (2018)
Xtriage (Phenix) : 1.13
EDS : trunk30967
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac : 5.8.0158
CCP4 : 7.0 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : trunk30967

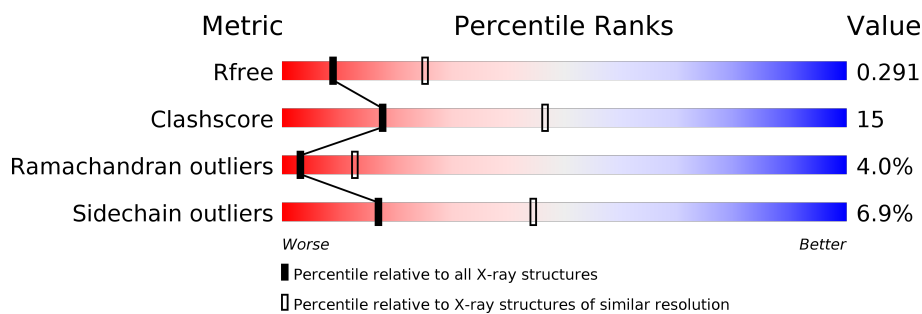
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.90 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}	111664	1716 (2.90-2.90)
Clashscore	122126	1924 (2.90-2.90)
Ramachandran outliers	120053	1884 (2.90-2.90)
Sidechain outliers	120020	1886 (2.90-2.90)

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. 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 $\leq 5\%$.

Mol	Chain	Length	Quality of chain
1	A	513	 77% 16% . .
1	B	513	 69% 21% 5% . .
1	C	513	 71% 22% . . .
1	D	513	 71% 21% 5% . .
1	E	513	 73% 18% 6% .
1	F	513	 74% 19% . .

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 24200 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.

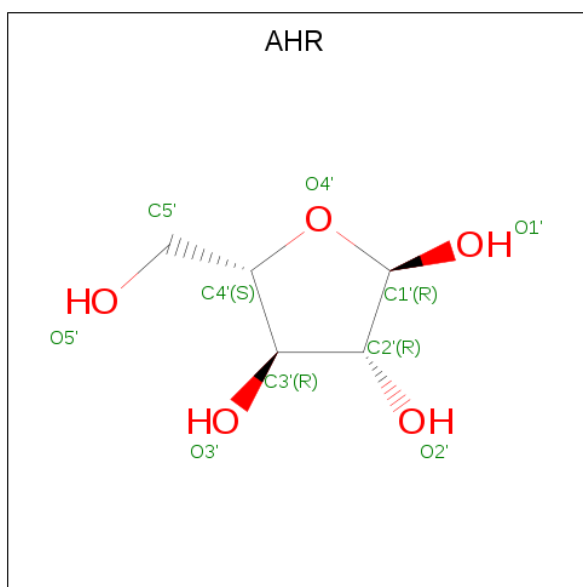
- Molecule 1 is a protein called ALPHA-L-ARABINOFURANOSIDASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	501	Total	C	N	O	S	0	0	0
			4017	2552	683	760	22			
1	B	500	Total	C	N	O	S	0	0	0
			4013	2545	686	760	22			
1	C	499	Total	C	N	O	S	0	1	0
			3989	2533	681	753	22			
1	D	498	Total	C	N	O	S	0	0	0
			3933	2487	672	753	21			
1	E	497	Total	C	N	O	S	0	0	0
			3927	2496	670	741	20			
1	F	500	Total	C	N	O	S	0	0	0
			3966	2518	679	748	21			

There are 6 discrepancies between the modelled and reference sequences:

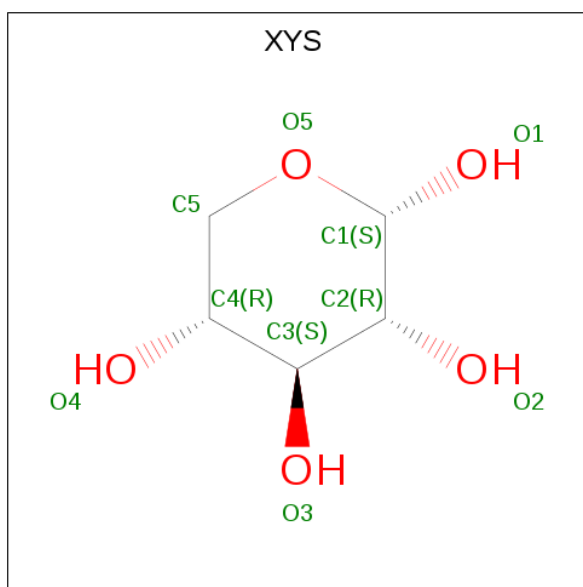
Chain	Residue	Modelled	Actual	Comment	Reference
A	173	ALA	GLU	engineered mutation	UNP Q4CJG5
B	173	ALA	GLU	engineered mutation	UNP Q4CJG5
C	173	ALA	GLU	engineered mutation	UNP Q4CJG5
D	173	ALA	GLU	engineered mutation	UNP Q4CJG5
E	173	ALA	GLU	engineered mutation	UNP Q4CJG5
F	173	ALA	GLU	engineered mutation	UNP Q4CJG5

- Molecule 2 is ALPHA-L-ARABINOFURANOSE (three-letter code: AHR) (formula: $C_5H_{10}O_5$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			9	5	4		
2	B	1	Total	C	O	0	0
			9	5	4		
2	C	1	Total	C	O	0	0
			9	5	4		
2	D	1	Total	C	O	0	0
			9	5	4		
2	E	1	Total	C	O	0	0
			9	5	4		
2	F	1	Total	C	O	0	0
			9	5	4		

- Molecule 3 is XYLOPYRANOSE (three-letter code: XYZ) (formula: $C_5H_{10}O_5$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			9	5	4		
3	B	1	Total	C	O	0	0
			9	5	4		
3	C	1	Total	C	O	0	0
			9	5	4		
3	D	1	Total	C	O	0	0
			9	5	4		
3	E	1	Total	C	O	0	0
			9	5	4		
3	F	1	Total	C	O	0	0
			9	5	4		

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C₂H₆O₂).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C O 4 2 2	0	0
4	A	1	Total C O 4 2 2	0	0
4	B	1	Total C O 4 2 2	0	0
4	C	1	Total C O 4 2 2	0	0
4	C	1	Total C O 4 2 2	0	0
4	D	1	Total C O 4 2 2	0	0
4	E	1	Total C O 4 2 2	0	0
4	E	1	Total C O 4 2 2	0	0
4	F	1	Total C O 4 2 2	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	35	Total O 35 35	0	0
5	B	37	Total O 37 37	0	0
5	C	36	Total O 36 36	0	0

Continued on next page...

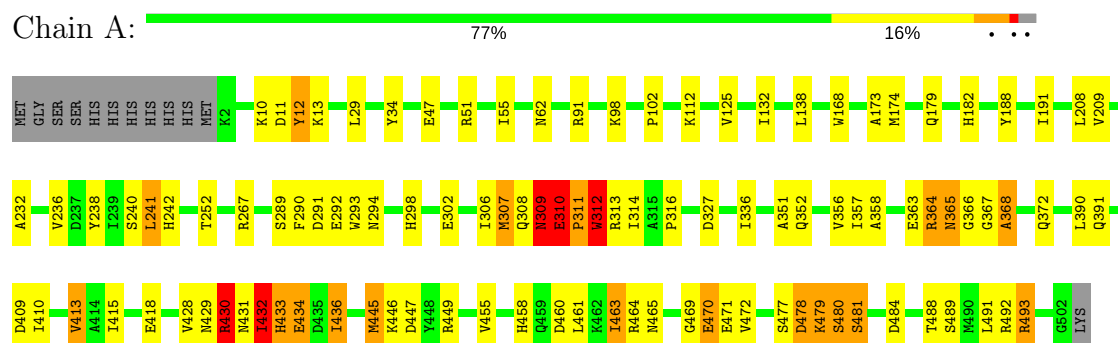
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	D	31	Total 31	O 31	0	0
5	E	40	Total 40	O 40	0	0
5	F	32	Total 32	O 32	0	0

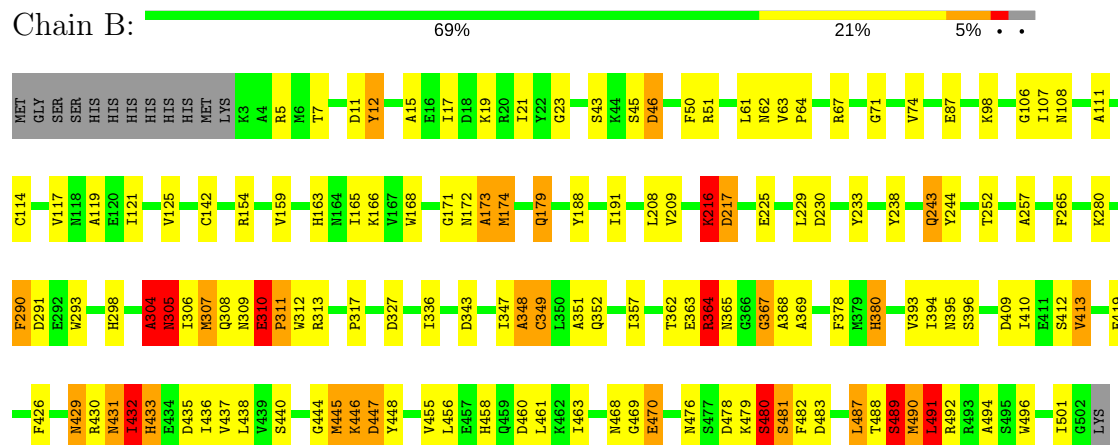
3 Residue-property plots

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. 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. 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.

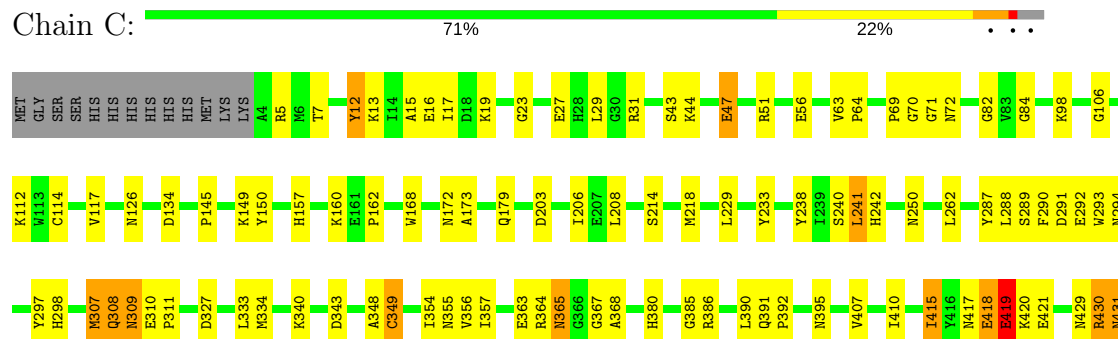
• Molecule 1: ALPHA-L-ARABINOFURANOSIDASE

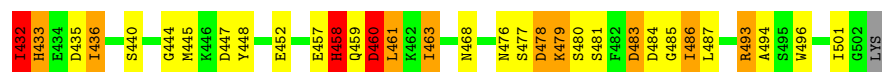


• Molecule 1: ALPHA-L-ARABINOFURANOSIDASE



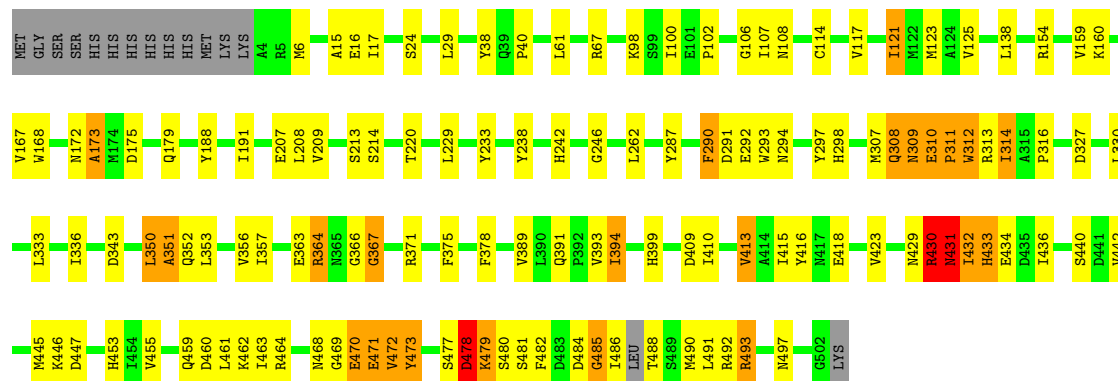
• Molecule 1: ALPHA-L-ARABINOFURANOSIDASE





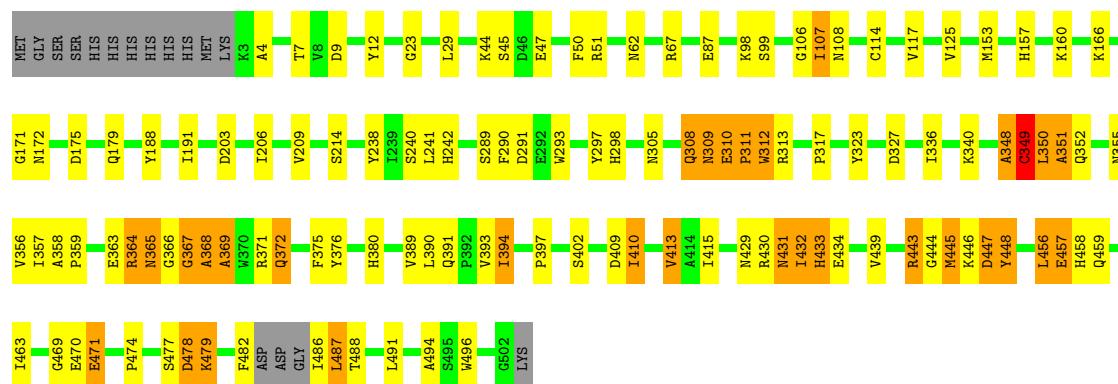
• Molecule 1: ALPHA-L-ARABINOFURANOSIDASE

Chain D: 71% 21% 5% • •



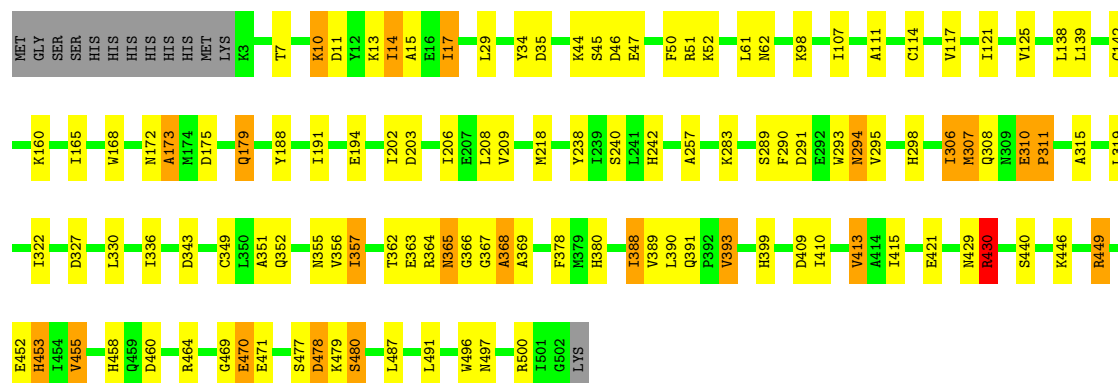
• Molecule 1: ALPHA-L-ARABINOFURANOSIDASE

Chain E: 73% 18% 6% •



• Molecule 1: ALPHA-L-ARABINOFURANOSIDASE

Chain F: 74% 19% • •



4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α , β , γ	173.45Å 173.45Å 272.20Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	145.86 – 2.90 63.35 – 2.43	Depositor EDS
% Data completeness (in resolution range)	98.8 (145.86-2.90) 92.4 (63.35-2.43)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.25 (at 2.42Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.201 , 0.255 0.249 , 0.291	Depositor DCC
R_{free} test set	7173 reflections (4.97%)	wwPDB-VP
Wilson B-factor (Å ²)	54.5	Xtriage
Anisotropy	0.066	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 53.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	24200	wwPDB-VP
Average B, all atoms (Å ²)	53.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.05% 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: AHR, XYS, EDO

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	0.38	0/4110	0.61	1/5566 (0.0%)
1	B	0.37	0/4105	0.60	0/5558
1	C	0.38	0/4082	0.60	0/5531
1	D	0.37	0/4021	0.59	0/5447
1	E	0.39	1/4019 (0.0%)	0.61	0/5447
1	F	0.37	0/4058	0.56	0/5496
All	All	0.38	1/24395 (0.0%)	0.60	1/33045 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	5
1	B	0	6
1	C	0	2
1	D	0	4
1	E	0	3
All	All	0	20

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	402	SER	CB-OG	5.30	1.49	1.42

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	310	GLU	C-N-CD	-9.15	100.47	120.60

There are no chirality outliers.

5 of 20 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	309	ASN	Peptide
1	A	310	GLU	Peptide
1	A	312	TRP	Peptide
1	A	432	ILE	Peptide
1	A	433	HIS	Peptide

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	4017	0	3920	117	0
1	B	4013	0	3908	158	0
1	C	3989	0	3870	107	0
1	D	3933	0	3767	113	0
1	E	3927	0	3786	109	0
1	F	3966	0	3845	90	0
2	A	9	0	8	0	0
2	B	9	0	8	1	0
2	C	9	0	8	0	0
2	D	9	0	8	0	0
2	E	9	0	8	0	0
2	F	9	0	8	0	0
3	A	9	0	7	0	0
3	B	9	0	7	0	0
3	C	9	0	7	0	0
3	D	9	0	7	0	0
3	E	9	0	7	0	0
3	F	9	0	7	0	0
4	A	8	0	12	0	0
4	B	4	0	6	0	0
4	C	8	0	12	0	0
4	D	4	0	6	0	0
4	E	8	0	12	0	0
4	F	4	0	6	0	0
5	A	35	0	0	1	0
5	B	37	0	0	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	C	36	0	0	0	0
5	D	31	0	0	0	0
5	E	40	0	0	1	0
5	F	32	0	0	1	0
All	All	24200	0	23240	691	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 15.

The worst 5 of 691 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:B:431:ASN:CB	1:B:432:ILE:HG22	1.11	1.56
1:D:431:ASN:HB3	1:D:432:ILE:C	1.38	1.40
1:B:431:ASN:CB	1:B:432:ILE:CG2	2.05	1.32
1:A:310:GLU:HB3	1:A:311:PRO:CB	1.73	1.18
1:E:478:ASP:HA	1:E:479:LYS:HB2	1.22	1.18

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	499/513 (97%)	445 (89%)	37 (7%)	17 (3%)	4	17
1	B	497/513 (97%)	435 (88%)	40 (8%)	22 (4%)	3	11
1	C	498/513 (97%)	446 (90%)	32 (6%)	20 (4%)	3	13
1	D	492/513 (96%)	428 (87%)	42 (8%)	22 (4%)	3	10
1	E	493/513 (96%)	436 (88%)	32 (6%)	25 (5%)	2	8
1	F	498/513 (97%)	445 (89%)	40 (8%)	13 (3%)	6	23

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
All	All	2977/3078 (97%)	2635 (88%)	223 (8%)	119 (4%)	3	13

5 of 119 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	62	ASN
1	A	309	ASN
1	A	310	GLU
1	A	311	PRO
1	A	364	ARG

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	434/450 (96%)	406 (94%)	28 (6%)	19	48
1	B	434/450 (96%)	398 (92%)	36 (8%)	12	35
1	C	428/450 (95%)	395 (92%)	33 (8%)	14	38
1	D	420/450 (93%)	393 (94%)	27 (6%)	19	49
1	E	417/450 (93%)	391 (94%)	26 (6%)	20	51
1	F	423/450 (94%)	396 (94%)	27 (6%)	19	49
All	All	2556/2700 (95%)	2379 (93%)	177 (7%)	17	44

5 of 177 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	307	MET
1	D	121	ILE
1	F	306	ILE
1	C	390	LEU
1	C	436	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 52 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	458	HIS
1	D	308	GLN
1	F	399	HIS
1	D	72	ASN
1	D	242	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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

21 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$
4	EDO	A	1503	-	3,3,3	0.56	0	2,2,2	0.16	0
4	EDO	A	1504	-	3,3,3	0.54	0	2,2,2	0.21	0
2	AHR	A	998	3	9,9,10	0.50	0	9,12,14	1.27	1 (11%)
3	XYS	A	999	2	9,9,10	0.84	0	10,12,14	1.89	2 (20%)
4	EDO	B	1503	-	3,3,3	0.58	0	2,2,2	0.12	0
2	AHR	B	998	3	9,9,10	0.55	0	9,12,14	1.60	2 (22%)
3	XYS	B	999	2	9,9,10	0.89	0	10,12,14	1.98	3 (30%)
4	EDO	C	1503	-	3,3,3	0.56	0	2,2,2	0.22	0
4	EDO	C	1504	-	3,3,3	0.49	0	2,2,2	0.37	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
2	AHR	C	998	3	9,9,10	0.43	0	9,12,14	1.01	1 (11%)
3	XYS	C	999	2	9,9,10	0.80	0	10,12,14	2.14	3 (30%)
4	EDO	D	1503	-	3,3,3	0.48	0	2,2,2	0.33	0
2	AHR	D	998	3	9,9,10	0.53	0	9,12,14	1.16	1 (11%)
3	XYS	D	999	2	9,9,10	0.94	0	10,12,14	2.15	3 (30%)
4	EDO	E	1503	-	3,3,3	0.66	0	2,2,2	0.22	0
4	EDO	E	1504	-	3,3,3	0.44	0	2,2,2	0.52	0
2	AHR	E	998	3	9,9,10	0.57	0	9,12,14	1.21	1 (11%)
3	XYS	E	999	2	9,9,10	0.87	0	10,12,14	1.66	2 (20%)
4	EDO	F	1503	-	3,3,3	0.48	0	2,2,2	0.27	0
2	AHR	F	998	3	9,9,10	0.51	0	9,12,14	1.06	1 (11%)
3	XYS	F	999	2	9,9,10	0.88	0	10,12,14	1.59	2 (20%)

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
4	EDO	A	1503	-	-	0/1/1/1	0/0/0/0
4	EDO	A	1504	-	-	0/1/1/1	0/0/0/0
2	AHR	A	998	3	-	0/2/15/18	0/1/1/1
3	XYS	A	999	2	-	0/0/14/17	1/1/1/1
4	EDO	B	1503	-	-	0/1/1/1	0/0/0/0
2	AHR	B	998	3	-	0/2/15/18	0/1/1/1
3	XYS	B	999	2	-	0/0/14/17	1/1/1/1
4	EDO	C	1503	-	-	0/1/1/1	0/0/0/0
4	EDO	C	1504	-	-	0/1/1/1	0/0/0/0
2	AHR	C	998	3	-	0/2/15/18	0/1/1/1
3	XYS	C	999	2	-	0/0/14/17	0/1/1/1
4	EDO	D	1503	-	-	0/1/1/1	0/0/0/0
2	AHR	D	998	3	-	0/2/15/18	0/1/1/1
3	XYS	D	999	2	-	0/0/14/17	1/1/1/1
4	EDO	E	1503	-	-	0/1/1/1	0/0/0/0
4	EDO	E	1504	-	-	0/1/1/1	0/0/0/0
2	AHR	E	998	3	-	0/2/15/18	0/1/1/1
3	XYS	E	999	2	-	0/0/14/17	0/1/1/1
4	EDO	F	1503	-	-	0/1/1/1	0/0/0/0
2	AHR	F	998	3	-	0/2/15/18	0/1/1/1
3	XYS	F	999	2	-	0/0/14/17	0/1/1/1

There are no bond length outliers.

The worst 5 of 22 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	D	999	XYS	C4-C3-C2	-2.51	107.94	110.86
3	C	999	XYS	C4-C3-C2	-2.51	107.94	110.86
2	B	998	AHR	C5'-C4'-C3'	-2.36	109.34	115.06
3	E	999	XYS	O4-C4-C5	-2.30	104.55	109.17
3	B	999	XYS	C4-C3-C2	-2.09	108.42	110.86

There are no chirality outliers.

There are no torsion outliers.

All (3) ring outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	B	999	XYS	C1-C2-C3-C4-C5-O5
3	A	999	XYS	C1-C2-C3-C4-C5-O5
3	D	999	XYS	C1-C2-C3-C4-C5-O5

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	998	AHR	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 ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.