WORLDWIDE PDB

PROTEIN DATA BANK

Full wwPDB X-ray Structure Validation Report (i)

Jun 23, 2024 – 02:38 PM EDT

PDB ID	:	4YYF
Title	:	The crystal structure of a glycosyl hydrolase of GH3 family member from [Mycobacterium smegmatis str. MC2 155
Authors	:	Tan, K.; Hatzos-Skintges, C.; Clancy, S.; Joachimiak, A.; Midwest Center for Structural Genomics (MCSG)
Deposited on	:	2015-03-23
Resolution	:	1.92 Å(reported)

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

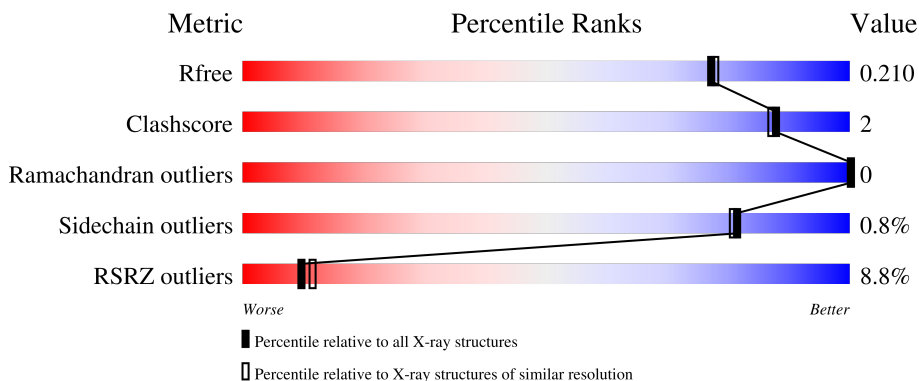|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 1.92 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	Percentile Ranks	Value
Rfree		0.210
Clashscore		2
Ramachandran outliers		0
Sidechain outliers		0.8%
RSRZ outliers		8.8%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	7937 (1.94-1.90)
Clashscore	141614	8644 (1.94-1.90)
Ramachandran outliers	138981	8530 (1.94-1.90)
Sidechain outliers	138945	8530 (1.94-1.90)
RSRZ outliers	127900	7793 (1.94-1.90)

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 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	365	% 88% • 9%
1	B	365	% 85% 5% 10%
1	C	365	22% 81% 7% 12%

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
2	ACT	B	403	-	-	-	X

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 7645 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 Beta-N-acetylhexosaminidase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	331	Total	C	N	O	S	Se	0	1	0
			2387	1481	427	470	2	7			
1	B	330	Total	C	N	O	S	Se	0	0	0
			2367	1469	422	468	1	7			
1	C	320	Total	C	N	O	S	Se	0	0	0
			2301	1430	412	451	2	6			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	23	SER	-	expression tag	UNP A0QPD6
A	24	ASN	-	expression tag	UNP A0QPD6
A	25	ALA	-	expression tag	UNP A0QPD6
A	291	VAL	ALA	engineered mutation	UNP A0QPD6
B	23	SER	-	expression tag	UNP A0QPD6
B	24	ASN	-	expression tag	UNP A0QPD6
B	25	ALA	-	expression tag	UNP A0QPD6
B	291	VAL	ALA	engineered mutation	UNP A0QPD6
C	23	SER	-	expression tag	UNP A0QPD6
C	24	ASN	-	expression tag	UNP A0QPD6
C	25	ALA	-	expression tag	UNP A0QPD6
C	291	VAL	ALA	engineered mutation	UNP A0QPD6

- Molecule 2 is ACETATE ION (three-letter code: ACT) (formula: $C_2H_3O_2$).

ACT

OXT

O⁻

C

CH3

O

O

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

- Molecule 3 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total 1	Na 1	0	0
3	B	2	Total 2	Na 2	0	0

- Molecule 4 is FORMIC ACID (three-letter code: FMT) (formula: CH_2O_2).

FMT

C

OH O2

O

O1

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total 3	C 1	O 2	0	0

- Molecule 5 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).

GOL

C3

OH O3

C1　　C2

HO

O1

OH

O2

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

- Molecule 6 is water.

W O R L D W I D E
PDB
PROTEIN DATA BANK

| Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	236	Total O 236 236	0	0
6	B	227	Total O 228 228	0	1
6	C	90	Total O 90 90	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.
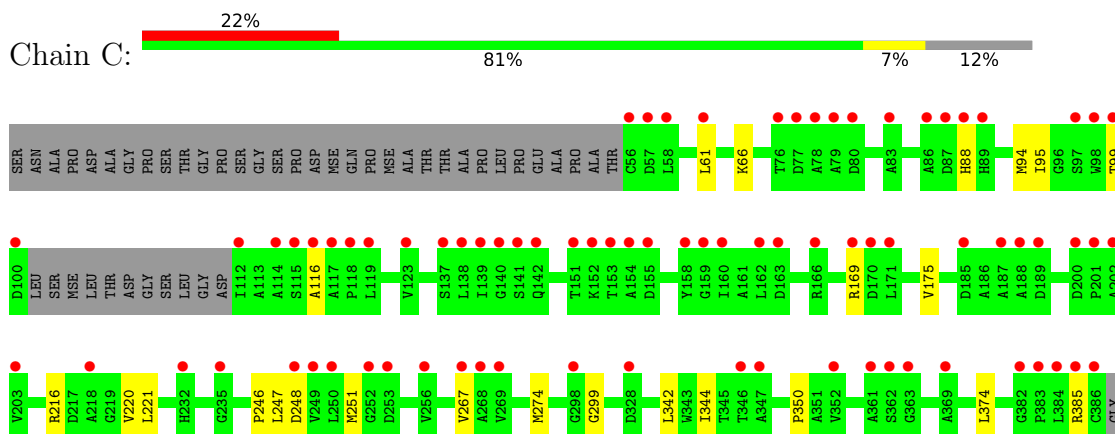
- Molecule 1: Beta-N-acetylhexosaminidase
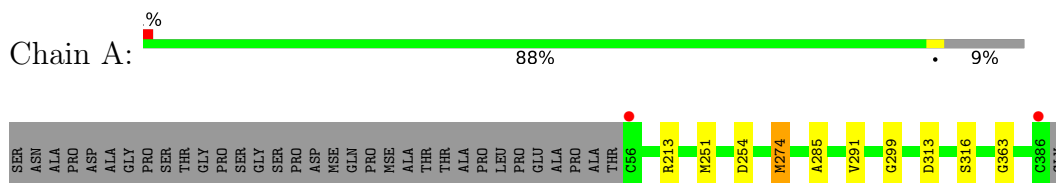
- Molecule 1: Beta-N-acetylhexosaminidase
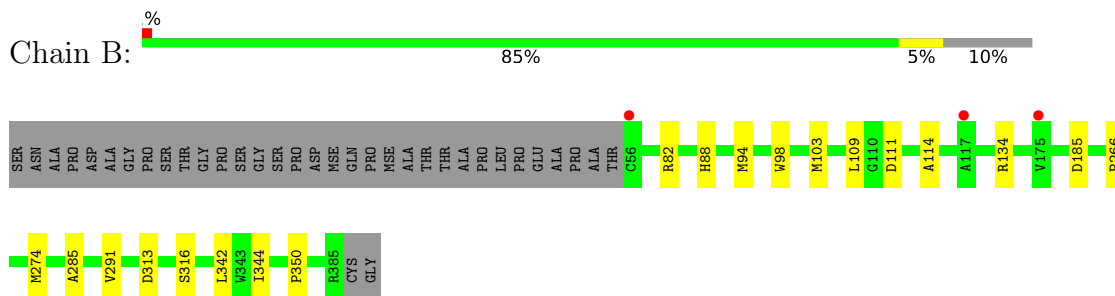
- Molecule 1: Beta-N-acetylhexosaminidase

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	103.07Å 130.94Å 85.99Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	36.41 – 1.92 36.41 – 1.92	Depositor EDS
% Data completeness (in resolution range)	99.0 (36.41-1.92) 99.1 (36.41-1.92)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.45 (at 1.92Å)	Xtriage
Refinement program	PHENIX 1.8.2_1309	Depositor
R, R_{free}	0.179 , 0.210 0.179 , 0.210	Depositor DCC
R_{free} test set	4376 reflections (4.94%)	wwPDB-VP
Wilson B-factor (Å^2)	20.1	Xtriage
Anisotropy	0.037	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 48.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.50, < L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	7645	wwPDB-VP
Average B, all atoms (Å^2)	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.88% 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: FMT, GOL, ACT, NA

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	0.37	0/2420	0.57	0/3290
1	B	0.37	0/2400	0.59	1/3264 (0.0%)
1	C	0.30	0/2334	0.49	0/3175
All	All	0.35	0/7154	0.55	1/9729 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	342	LEU	CA-CB-CG	5.34	127.59	115.30

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	2387	0	2365	6	0
1	B	2367	0	2340	10	0
1	C	2301	0	2280	13	0
2	A	8	0	6	0	0
2	B	12	0	9	0	0
2	C	4	0	3	0	0
3	A	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	2	0	0	0	0
4	A	3	0	1	0	0
5	B	6	0	8	0	0
6	A	236	0	0	1	0
6	B	228	0	0	2	0
6	C	90	0	0	0	1
All	All	7645	0	7012	28	1

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 2.

All (28) 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:313:ASP:OD2	1:B:316:SER:OG	2.23	0.56
1:C:88:HIS:NE2	1:C:350:PRO:HG3	2.23	0.54
1:A:213:ARG:NH2	6:A:505:HOH:O	2.40	0.53
1:C:246:PRO:HB2	1:C:248:ASP:OD1	2.10	0.52
1:C:247:LEU:HD11	1:C:251:MSE:HE2	1.92	0.52
1:B:82:ARG:NH1	6:B:503:HOH:O	2.33	0.51
1:B:94:MSE:HE1	1:B:344:ILE:HG22	1.92	0.51
1:B:285:ALA:O	1:B:291:VAL:HG21	2.12	0.50
1:A:363:GLY:HA3	1:B:114:ALA:HB2	1.92	0.50
1:C:66:LYS:HG2	1:C:374:LEU:HD22	1.94	0.49
1:C:251:MSE:HE1	1:C:299:GLY:O	2.11	0.49
1:C:94:MSE:HE2	1:C:342:LEU:HD21	1.95	0.49
1:A:285:ALA:O	1:A:291:VAL:HG21	2.14	0.48
1:B:88:HIS:CE1	1:B:350:PRO:HG3	2.49	0.47
1:C:95:ILE:HG23	1:C:99:THR:HG21	1.96	0.47
1:C:94:MSE:HE1	1:C:344:ILE:HG22	1.96	0.46
1:C:116:ALA:O	1:C:385:ARG:NH1	2.48	0.46
1:A:251:MSE:HE1	1:A:299:GLY:O	2.18	0.43
1:B:82:ARG:NH2	1:B:111:ASP:OD2	2.48	0.43
1:B:103:MSE:HE2	1:B:109:LEU:HD13	2.00	0.43
1:C:61:LEU:O	1:C:66:LYS:HE3	2.19	0.43
1:B:94:MSE:HE3	1:B:98:TRP:CZ2	2.54	0.42
1:C:175:VAL:HG22	1:C:221:LEU:HB3	2.01	0.41
1:A:313:ASP:OD2	1:A:316:SER:OG	2.30	0.41
1:C:169:ARG:HG2	1:C:220:VAL:CG2	2.50	0.41
1:B:266:PRO:HG2	6:B:698[A]:HOH:O	2.21	0.41
1:A:274:MSE:SE	1:A:274:MSE:H	2.54	0.41

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:216:ARG:HH11	1:C:267:VAL:HB	1.87	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
6:C:588:HOH:O	6:C:588:HOH:O[2_755]	2.08	0.12

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	330/365 (90%)	322 (98%)	8 (2%)	0	100	100
1	B	328/365 (90%)	320 (98%)	8 (2%)	0	100	100
1	C	316/365 (87%)	308 (98%)	8 (2%)	0	100	100
All	All	974/1095 (89%)	950 (98%)	24 (2%)	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	244/258 (95%)	242 (99%)	2 (1%)	81	81

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	241/258 (93%)	238 (99%)	3 (1%)	71	69
1	C	234/258 (91%)	233 (100%)	1 (0%)	91	91
All	All	719/774 (93%)	713 (99%)	6 (1%)	81	81

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	254	ASP
1	A	274	MSE
1	B	134	ARG
1	B	185	ASP
1	B	274	MSE
1	C	274	MSE

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

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 ⓘ

Of 11 ligands modelled in this entry, 3 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	#$|Z| > 2$
2	ACT	C	401	-	3,3,3	0.74	0	3,3,3	1.42	0
5	GOL	B	401	-	5,5,5	0.26	0	5,5,5	0.40	0
4	FMT	A	404	-	2,2,2	0.74	0	1,1,1	0.25	0
2	ACT	A	402	-	3,3,3	0.78	0	3,3,3	1.45	0
2	ACT	A	401	-	3,3,3	0.80	0	3,3,3	1.30	0
2	ACT	B	404	-	3,3,3	0.77	0	3,3,3	1.39	0
2	ACT	B	403	-	3,3,3	0.75	0	3,3,3	1.34	0
2	ACT	B	402	-	3,3,3	0.74	0	3,3,3	1.30	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	GOL	B	401	-	-	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 ⓘ

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	324/365 (88%)	-0.24	2 (0%)	89	90	13, 20, 35, 71	0
1	B	323/365 (88%)	-0.16	3 (0%)	84	85	12, 20, 34, 78	0
1	C	314/365 (86%)	1.47	80 (25%)	0	0	36, 46, 66, 101	0
All	All	961/1095 (87%)	0.35	85 (8%)	10	11	12, 25, 57, 101	0

All (85) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	386	CYS	8.9
1	C	56	CYS	7.4
1	C	117	ALA	6.9
1	A	386	CYS	5.3
1	C	138	LEU	5.0
1	C	163	ASP	4.5
1	C	98	TRP	4.5
1	C	77	ASP	4.3
1	C	155	ASP	4.3
1	B	56	CYS	4.2
1	C	79	ALA	4.1
1	C	139	ILE	4.1
1	C	154	ALA	4.0
1	C	170	ASP	4.0
1	C	116	ALA	4.0
1	C	384	LEU	3.9
1	A	56	CYS	3.9
1	C	383	PRO	3.9
1	C	83	ALA	3.8
1	C	114	ALA	3.8
1	C	140	GLY	3.7
1	C	169	ARG	3.7
1	C	363	GLY	3.5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	C	256	VAL	3.5
1	C	137	SER	3.4
1	C	202	ALA	3.4
1	C	80	ASP	3.3
1	C	141	SER	3.3
1	C	249	VAL	3.3
1	C	57	ASP	3.3
1	C	162	LEU	3.2
1	C	112	ILE	3.2
1	C	115	SER	3.2
1	C	218	ALA	3.1
1	C	252	GLY	3.1
1	C	151	THR	3.1
1	C	385	ARG	3.1
1	C	268	ALA	3.1
1	B	117	ALA	3.0
1	C	188	ALA	3.0
1	C	361	ALA	2.9
1	C	158	TYR	2.9
1	C	118	PRO	2.9
1	C	159	GLY	2.8
1	C	99	THR	2.8
1	C	235	GLY	2.8
1	C	201	PRO	2.7
1	C	166	ARG	2.7
1	C	346	THR	2.7
1	C	382	GLY	2.7
1	C	362	SER	2.7
1	C	89	HIS	2.6
1	C	153	THR	2.6
1	C	189	ASP	2.6
1	C	97	SER	2.6
1	C	250	LEU	2.6
1	C	187	ALA	2.6
1	C	253	ASP	2.6
1	C	267	VAL	2.5
1	C	298	GLY	2.4
1	C	171	LEU	2.4
1	C	269	VAL	2.4
1	C	86	ALA	2.4
1	C	347	ALA	2.4
1	C	248	ASP	2.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	C	87	ASP	2.3
1	C	119	LEU	2.3
1	C	152	LYS	2.3
1	C	203	VAL	2.3
1	C	352	VAL	2.3
1	C	88	HIS	2.3
1	C	100	ASP	2.3
1	C	328	ASP	2.2
1	C	58	LEU	2.2
1	C	61	LEU	2.2
1	C	76	THR	2.2
1	C	200	ASP	2.2
1	C	160	ILE	2.2
1	C	369	ALA	2.2
1	C	232	HIS	2.1
1	C	123	VAL	2.1
1	C	185	ASP	2.1
1	C	78	ALA	2.1
1	C	142	GLN	2.0
1	B	175	VAL	2.0

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
2	ACT	B	403	4/4	0.79	0.46	49,53,57,65	0
4	FMT	A	404	3/3	0.85	0.20	53,53,54,55	0
2	ACT	B	404	4/4	0.90	0.46	54,55,57,60	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	ACT	A	402	4/4	0.91	0.17	28,30,35,38	0
3	NA	B	405	1/1	0.95	0.08	33,33,33,33	0
2	ACT	C	401	4/4	0.96	0.16	31,32,32,33	0
5	GOL	B	401	6/6	0.96	0.10	23,32,33,33	0
2	ACT	A	401	4/4	0.98	0.09	21,21,21,22	0
2	ACT	B	402	4/4	0.98	0.15	18,21,21,23	0
3	NA	A	403	1/1	0.99	0.06	20,20,20,20	0
3	NA	B	406	1/1	0.99	0.06	21,21,21,21	0

6.5 Other polymers ⓘ

There are no such residues in this entry.