Full wwPDB X-ray Structure Validation Report ⓘ

Oct 31, 2023 – 06:39 PM JST

PDB ID	:	5GT8
Title	:	Crystal Structure of apo-CASTOR1
Authors	:	Guo, L.; Deng, D.
Deposited on	:	2016-08-18
Resolution	:	2.80 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

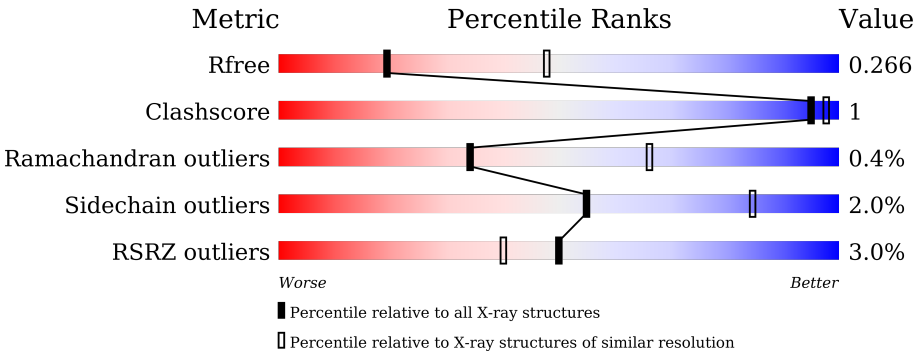
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 2.80 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	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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	335	2% 78% • 17%
1	B	335	% 76% 5% 19%
1	C	335	4% 76% • 20%
2	D	335	2% 74% 5% 21%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 16411 atoms, of which 8129 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 GATS-like protein 3.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	277	Total	C	H	N	O	S	0	1	0
			4253	1386	2126	348	386	7			
1	B	270	Total	C	H	N	O	S	0	0	0
			4109	1345	2046	331	381	6			
1	C	268	Total	C	H	N	O	S	0	0	0
			4022	1323	1987	332	373	7			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-5	ALA	-	expression tag	UNP Q8WTX7
A	-4	GLY	-	expression tag	UNP Q8WTX7
A	-3	SER	-	expression tag	UNP Q8WTX7
A	-2	GLY	-	expression tag	UNP Q8WTX7
A	-1	HIS	-	expression tag	UNP Q8WTX7
A	0	MET	-	expression tag	UNP Q8WTX7
B	-5	ALA	-	expression tag	UNP Q8WTX7
B	-4	GLY	-	expression tag	UNP Q8WTX7
B	-3	SER	-	expression tag	UNP Q8WTX7
B	-2	GLY	-	expression tag	UNP Q8WTX7
B	-1	HIS	-	expression tag	UNP Q8WTX7
B	0	MET	-	expression tag	UNP Q8WTX7
C	-5	ALA	-	expression tag	UNP Q8WTX7
C	-4	GLY	-	expression tag	UNP Q8WTX7
C	-3	SER	-	expression tag	UNP Q8WTX7
C	-2	GLY	-	expression tag	UNP Q8WTX7
C	-1	HIS	-	expression tag	UNP Q8WTX7
C	0	MET	-	expression tag	UNP Q8WTX7

- Molecule 2 is a protein called GATS-like protein 3.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	D	265	Total	C	H	N	O	S	0	0	0
			3983	1311	1970	327	368	7			

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	-5	ALA	-	expression tag	UNP Q8WTX7
D	-4	GLY	-	expression tag	UNP Q8WTX7
D	-3	SER	-	expression tag	UNP Q8WTX7
D	-2	GLY	-	expression tag	UNP Q8WTX7
D	-1	HIS	-	expression tag	UNP Q8WTX7
D	0	MET	-	expression tag	UNP Q8WTX7
D	285	ILE	ALA	engineered mutation	UNP Q8WTX7

- Molecule 3 is water.

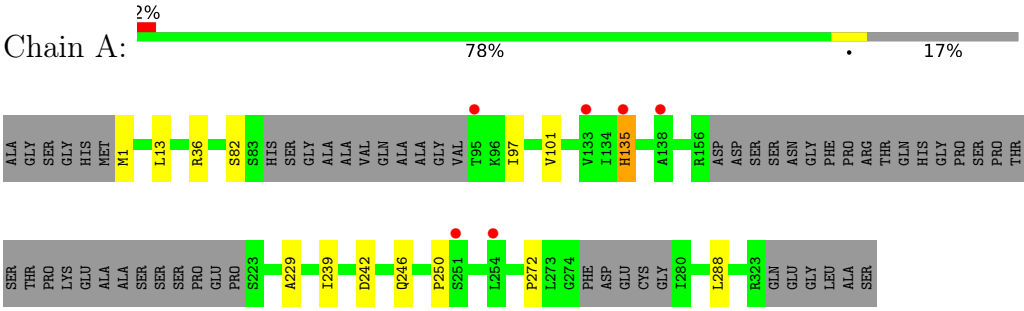
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	16	Total	O	0	0
			16	16		
3	B	12	Total	O	0	0
			12	12		
3	C	8	Total	O	0	0
			8	8		
3	D	8	Total	O	0	0
			8	8		

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.

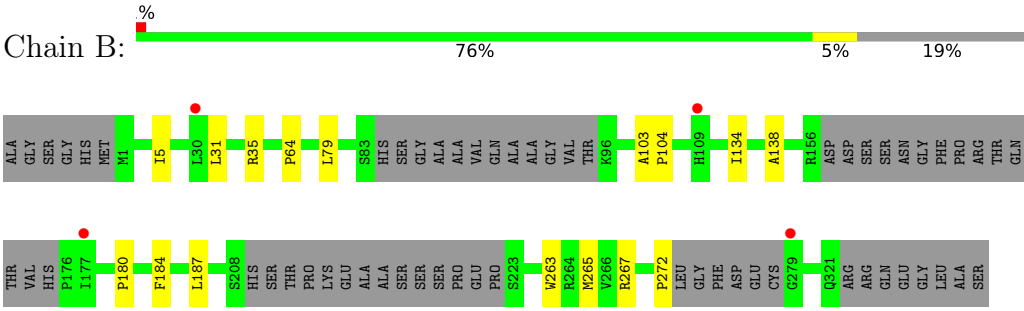
• Molecule 1: GATS-like protein 3

Chain A:



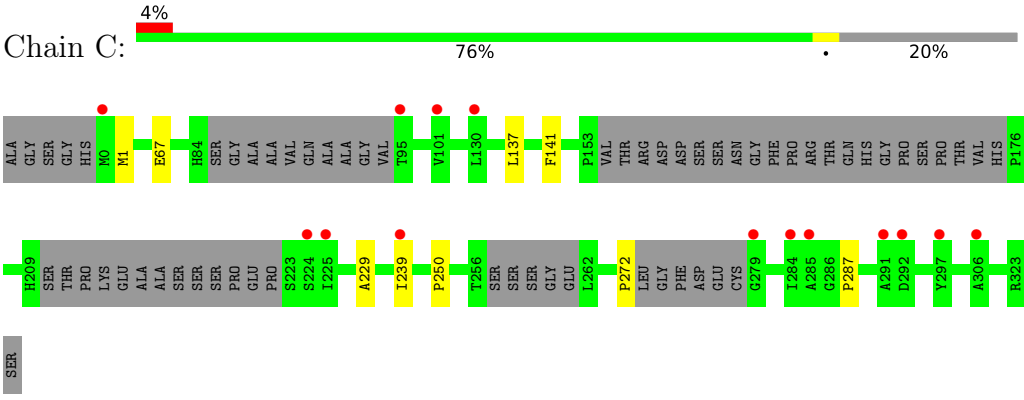
• Molecule 1: GATS-like protein 3

Chain B:



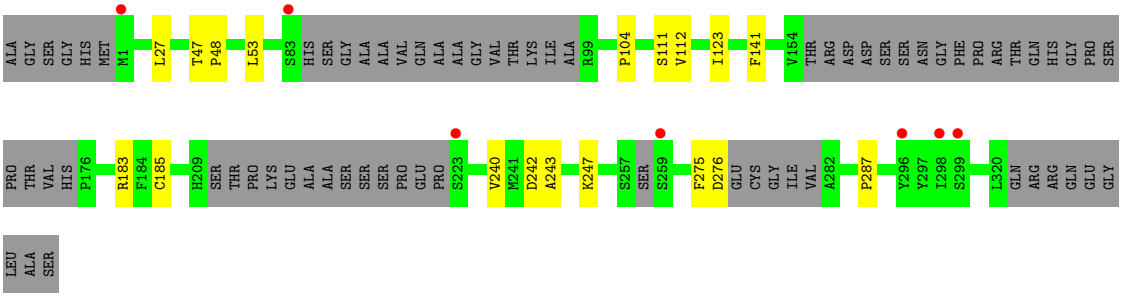
• Molecule 1: GATS-like protein 3

Chain C:



• Molecule 2: GATS-like protein 3

Chain D:


ALA GLY SER GLY HIS MET M1 L27 T47 P48 L53 S83 HIS SER GLY ALA ALA VAL GLN ALA ALA GLY VAL THR LYS ILE ALA R99 P104 S111 V112 I123 F141 V154 THR ARG ASP SER SER ASN GLY PHE PRO ARG THR GLN HIS GLY PRO SER
PRO THR VAL HIS P176 R183 F184 C185 H209 SER THR PRO LYS GLU ALA ALA SER SER SER PRO GLU PRO S223 V240 M241 D242 A243 K247 S257 SER S259 F275 D276 GLU CYS GLY ILE VAL A282 P287 Y296 Y297 I298 S299 L320 GLN ARG ARG GLN GLU GLY
LEU ALA SER

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	47.41Å 142.60Å 95.40Å 90.00° 90.26° 90.00°	Depositor
Resolution (Å)	47.41 – 2.80 95.40 – 2.70	Depositor EDS
% Data completeness (in resolution range)	99.6 (47.41-2.80) 92.4 (95.40-2.70)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.16 (at 2.69Å)	Xtriage
Refinement program	PHENIX 1.10.1_2155: ???	Depositor
R, R_{free}	0.243 , 0.265 0.244 , 0.266	Depositor DCC
R_{free} test set	2002 reflections (5.78%)	wwPDB-VP
Wilson B-factor (Å^2)	49.6	Xtriage
Anisotropy	0.826	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 40.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.44$, $< L^2 > = 0.27$	Xtriage
Estimated twinning fraction	0.156 for h,-k,-l	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	16411	wwPDB-VP
Average B, all atoms (Å^2)	73.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 17.91% 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

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	0.26	0/2177	0.47	1/2969 (0.0%)
1	B	0.26	0/2109	0.46	1/2878 (0.0%)
1	C	0.26	0/2081	0.48	2/2839 (0.1%)
2	D	0.26	0/2058	0.48	2/2805 (0.1%)
All	All	0.26	0/8425	0.47	6/11491 (0.1%)

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	C	0	1

There are no bond length outliers.

All (6) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	272	PRO	N-CA-CB	5.98	110.48	103.30
1	A	272	PRO	N-CA-CB	5.95	110.44	103.30
1	C	272	PRO	N-CA-CB	5.80	110.26	103.30
1	C	287	PRO	N-CA-CB	5.71	110.16	103.30
2	D	287	PRO	N-CA-CB	5.56	109.98	103.30
2	D	104	PRO	N-CA-CB	5.33	109.69	103.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	177	ILE	Peptide

5.2 Too-close contacts (i)

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	2127	2126	2087	5	0
1	B	2063	2046	2008	5	0
1	C	2035	1987	1935	3	0
2	D	2013	1970	1911	8	0
3	A	16	0	0	0	0
3	B	12	0	0	0	0
3	C	8	0	0	0	0
3	D	8	0	0	0	0
All	All	8282	8129	7941	20	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 1.

All (20) 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)
2:D:27:LEU:HD21	2:D:53:LEU:HD21	1.92	0.52
2:D:185:CYS:O	2:D:240:VAL:HA	2.12	0.49
1:A:36:ARG:NH1	1:C:203:ASP:OD1	2.47	0.47
2:D:275:PHE:O	2:D:276:ASP:CB	2.63	0.46
1:A:1:MET:N	1:A:82:SER:O	2.46	0.46
1:B:103:ALA:HB3	1:B:104:PRO:HD3	1.98	0.45
2:D:275:PHE:O	2:D:276:ASP:HB2	2.17	0.44
2:D:183:ARG:O	2:D:243:ALA:N	2.52	0.43
1:A:135:HIS:O	1:A:135:HIS:ND1	2.51	0.43
1:B:184:PHE:N	1:B:263:TRP:O	2.48	0.42
2:D:112:VAL:CG1	2:D:123:ILE:HG23	2.50	0.42
1:B:5:ILE:HG12	1:B:79:LEU:CD2	2.50	0.42
1:B:265:MET:HE3	1:B:267:ARG:HB3	2.02	0.41
2:D:112:VAL:HG13	2:D:123:ILE:HG23	2.01	0.41
1:B:134:ILE:O	1:B:138:ALA:N	2.53	0.41
2:D:47:THR:HB	2:D:48:PRO:HD2	2.01	0.41
1:A:97:ILE:HG22	1:A:101:VAL:HB	2.03	0.41
1:A:229:ALA:O	1:A:239:ILE:HA	2.21	0.41
1:C:1:MET:HB2	1:C:141:PHE:HD1	1.86	0.41

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:229:ALA:O	1:C:239:ILE:HA	2.21	0.41

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	268/335 (80%)	252 (94%)	15 (6%)	1 (0%)	34	66
1	B	260/335 (78%)	242 (93%)	17 (6%)	1 (0%)	34	66
1	C	256/335 (76%)	240 (94%)	14 (6%)	2 (1%)	19	49
2	D	253/335 (76%)	234 (92%)	19 (8%)	0	100	100
All	All	1037/1340 (77%)	968 (93%)	65 (6%)	4 (0%)	34	66

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	250	PRO
1	C	250	PRO
1	C	177	ILE
1	B	180	PRO

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	224/289 (78%)	218 (97%)	6 (3%)	44 78
1	B	217/289 (75%)	213 (98%)	4 (2%)	59 86
1	C	207/289 (72%)	204 (99%)	3 (1%)	67 90
2	D	203/290 (70%)	199 (98%)	4 (2%)	55 84
All	All	851/1157 (74%)	834 (98%)	17 (2%)	55 84

All (17) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	13	LEU
1	A	135	HIS
1	A	182	ASN
1	A	242	ASP
1	A	246	GLN
1	A	288	LEU
1	B	31	LEU
1	B	35	ARG
1	B	64	PRO
1	B	187	LEU
1	C	67	GLU
1	C	137	LEU
1	C	187	LEU
2	D	111	SER
2	D	141	PHE
2	D	242	ASP
2	D	247	LYS

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

Mol	Chain	Res	Type
1	B	246	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)

There are no ligands in this entry.

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	277/335 (82%)	0.11	6 (2%)	62	52	30, 58, 86, 100	0
1	B	270/335 (80%)	0.15	4 (1%)	73	68	33, 62, 85, 100	0
1	C	268/335 (80%)	0.39	15 (5%)	24	16	45, 79, 118, 138	0
2	D	265/335 (79%)	0.22	7 (2%)	56	46	36, 70, 100, 178	0
All	All	1080/1340 (80%)	0.22	32 (2%)	50	40	30, 66, 102, 178	0

All (32) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	95	THR	5.4
1	C	177	ILE	3.8
2	D	296	TYR	3.5
1	C	291	ALA	3.4
1	C	297	TYR	3.1
1	C	0	MET	3.1
1	C	292	ASP	3.1
2	D	83	SER	3.0
1	C	224	SER	3.0
1	B	30	LEU	3.0
1	A	95	THR	3.0
1	C	101	VAL	2.9
2	D	299	SER	2.9
1	C	279	GLY	2.9
1	C	285	ALA	2.9
2	D	1	MET	2.8
1	C	284	ILE	2.7
1	A	251	SER	2.6
2	D	259	SER	2.6
1	B	279	GLY	2.5
1	A	138	ALA	2.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	177	ILE	2.4
1	C	239	ILE	2.4
1	C	306	ALA	2.2
2	D	298	ILE	2.2
1	C	130	LEU	2.2
2	D	223	SER	2.2
1	C	225	ILE	2.1
1	A	254	LEU	2.1
1	A	135	HIS	2.0
1	B	109	HIS	2.0
1	A	133	VAL	2.0

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

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

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.