



Full wwPDB X-ray Structure Validation Report ⓘ

Mar 9, 2026 – 02:16 PM UTC

PDB ID : 4DVY / pdb_00004dvY
Title : Crystal structure of the Helicobacter pylori CagA oncoprotein
Authors : Hayashi, T.; Senda, M.; Morohashi, H.; Higashi, H.; Horio, M.; Kashiba, Y.;
Nagase, L.; Sasaya, D.; Shimizu, T.; Venugopalan, N.; Kumeta, H.; Noda, N.;
Inagaki, F.; Senda, T.; Hatakeyama, M.
Deposited on : 2012-02-23
Resolution : 3.30 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	:	4-5-2 with Phenix2.0
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

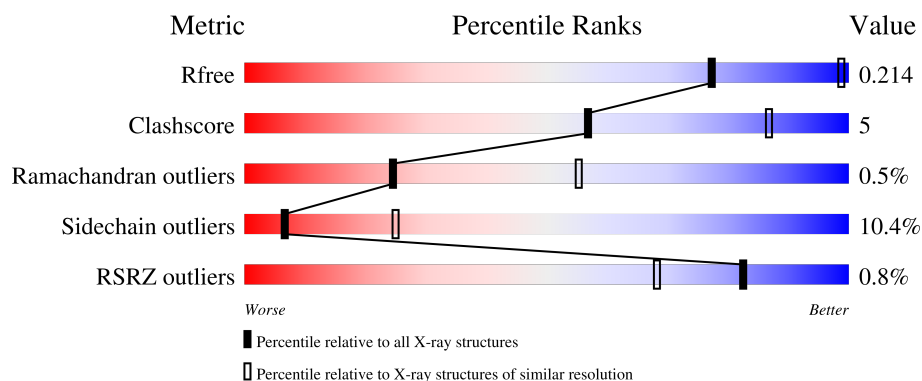
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

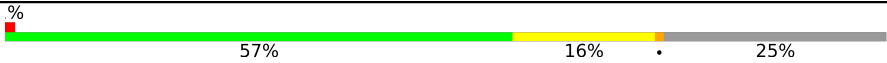
The reported resolution of this entry is 3.30 Å.

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}	180053	1169 (3.32-3.28)
Clashscore	190562	1209 (3.32-3.28)
Ramachandran outliers	187476	1188 (3.32-3.28)
Sidechain outliers	187428	1187 (3.32-3.28)
RSRZ outliers	180081	1169 (3.32-3.28)

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 $\leq 5\%$. 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	P	876	

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 5229 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 Cytotoxicity-associated immunodominant antigen.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	P	656	Total	C	N	O	S	0	0	0
			5229	3288	902	1030	9			

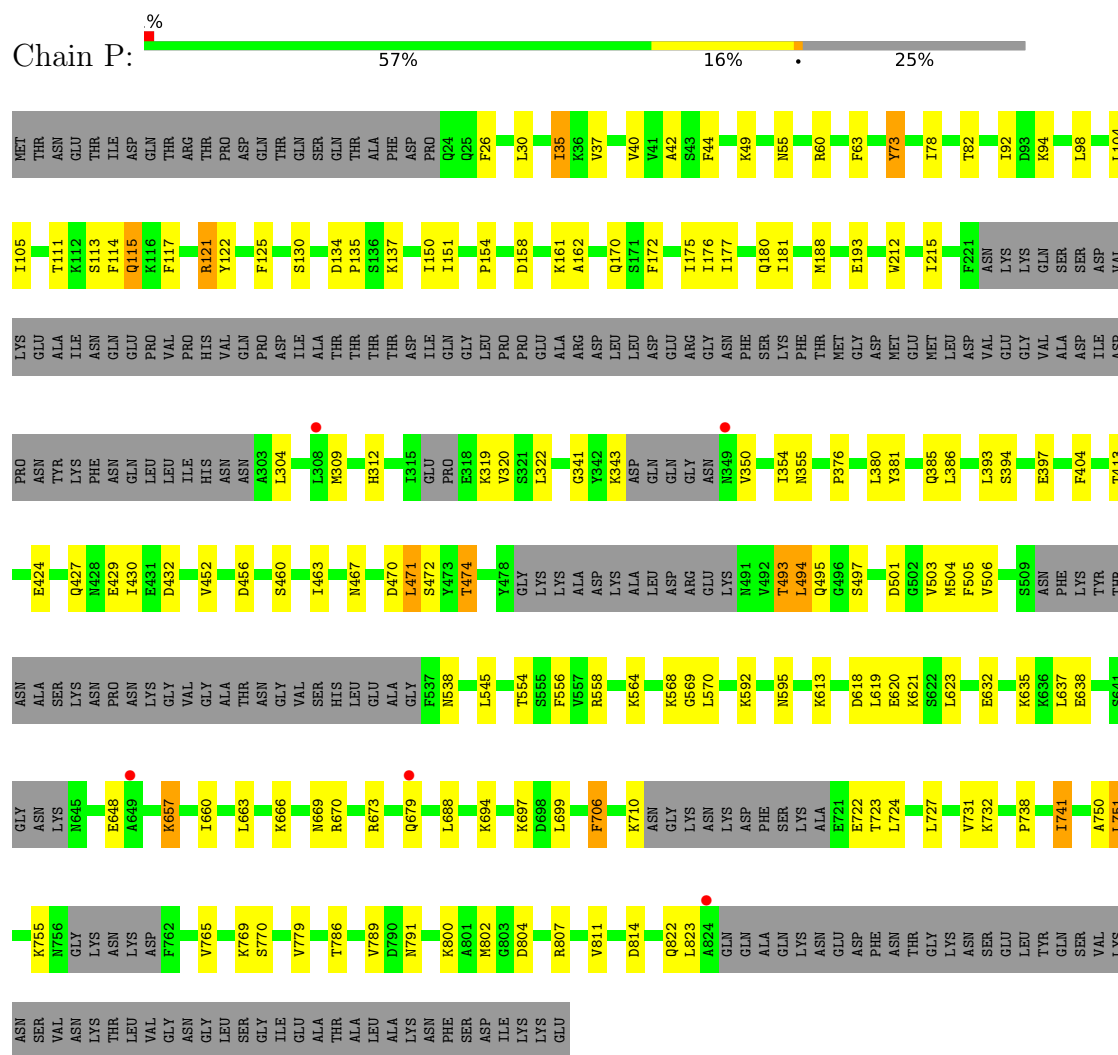
There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
P	131	MET	HIS	engineered mutation	UNP P55980
P	274	MET	LEU	engineered mutation	UNP P55980

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: Cytotoxicity-associated immunodominant antigen



4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α , β , γ	97.31Å 97.31Å 244.76Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	90.42 – 3.30 90.42 – 3.30	Depositor EDS
% Data completeness (in resolution range)	99.5 (90.42-3.30) 99.5 (90.42-3.30)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.69 (at 3.33Å)	Xtriage
Refinement program	BUSTER 2.10.0	Depositor
R, R_{free}	0.191 , 0.248 0.208 , 0.214	Depositor DCC
R_{free} test set	921 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	119.0	Xtriage
Anisotropy	0.020	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 151.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	5229	wwPDB-VP
Average B, all atoms (Å ²)	145.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.85% 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

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	P	0.83	1/5297 (0.0%)	1.52	37/7091 (0.5%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	P	802	MET	SD-CE	-6.51	1.63	1.79

All (37) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	P	755	LYS	N-CA-C	-8.27	104.33	114.75
1	P	122	TYR	N-CA-C	-7.15	103.57	111.36
1	P	538	ASN	N-CA-C	6.55	120.06	110.17
1	P	556	PHE	N-CA-C	-6.45	104.17	111.14
1	P	125	PHE	CA-CB-CG	6.37	120.17	113.80
1	P	121	ARG	CA-C-N	6.20	129.09	120.29
1	P	121	ARG	C-N-CA	6.20	129.09	120.29
1	P	55	ASN	CA-C-N	5.96	128.59	120.54
1	P	55	ASN	C-N-CA	5.96	128.59	120.54
1	P	470	ASP	N-CA-C	5.82	117.78	108.41
1	P	564	LYS	CA-C-N	5.80	128.05	120.28
1	P	564	LYS	C-N-CA	5.80	128.05	120.28
1	P	456	ASP	CA-CB-CG	5.78	118.38	112.60
1	P	137	LYS	N-CA-C	-5.71	105.20	111.82
1	P	193	GLU	CA-C-N	5.63	128.09	120.38
1	P	193	GLU	C-N-CA	5.63	128.09	120.38
1	P	497	SER	N-CA-C	5.63	117.95	107.99
1	P	55	ASN	CA-CB-CG	5.50	118.10	112.60
1	P	569	GLY	CA-C-N	5.47	131.99	121.54
1	P	569	GLY	C-N-CA	5.47	131.99	121.54
1	P	804	ASP	N-CA-C	5.41	116.24	107.32
1	P	130	SER	CA-C-N	5.41	128.30	120.79
1	P	130	SER	C-N-CA	5.41	128.30	120.79

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	P	413	THR	CB-CA-C	5.31	118.85	111.63
1	P	424	GLU	CA-C-N	5.29	127.63	120.65
1	P	424	GLU	C-N-CA	5.29	127.63	120.65
1	P	471	LEU	N-CA-C	5.25	118.11	109.76
1	P	638	GLU	CA-C-N	5.18	127.22	120.28
1	P	638	GLU	C-N-CA	5.18	127.22	120.28
1	P	73	TYR	N-CA-C	5.12	116.67	111.14
1	P	637	LEU	CA-C-N	5.07	127.07	120.28
1	P	637	LEU	C-N-CA	5.07	127.07	120.28
1	P	538	ASN	CA-C-N	5.06	130.27	122.93
1	P	538	ASN	C-N-CA	5.06	130.27	122.93
1	P	432	ASP	CA-C-N	5.05	127.01	120.44
1	P	432	ASP	C-N-CA	5.05	127.01	120.44
1	P	121	ARG	N-CA-C	-5.03	106.92	113.16

There are no chirality outliers.

There are no planarity outliers.

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	P	5229	0	5219	48	0
All	All	5229	0	5219	48	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.

All (48) 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:P:731:VAL:HG23	1:P:741:ILE:HG22	1.73	0.69
1:P:309:MET:HE1	1:P:381:TYR:CE1	2.30	0.66
1:P:60:ARG:HA	1:P:63:PHE:HD2	1.62	0.65
1:P:354:ILE:HD11	1:P:376:PRO:HB3	1.80	0.64
1:P:504:MET:HE1	1:P:506:VAL:HG22	1.82	0.60

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:P:135:PRO:HD2	1:P:427:GLN:HE22	1.66	0.60
1:P:309:MET:HE1	1:P:381:TYR:HE1	1.65	0.59
1:P:113:SER:HA	1:P:117:PHE:HB2	1.86	0.58
1:P:42:ALA:HB2	1:P:49:LYS:HG3	1.87	0.56
1:P:212:TRP:HA	1:P:215:ILE:HD12	1.87	0.56
1:P:172:PHE:HA	1:P:175:ILE:HD12	1.88	0.56
1:P:394:SER:OG	1:P:397:GLU:HG3	2.06	0.55
1:P:632:GLU:HA	1:P:635:LYS:HD2	1.88	0.55
1:P:495:GLN:HG3	1:P:505:PHE:HB2	1.89	0.55
1:P:37:VAL:HA	1:P:40:VAL:HG22	1.88	0.54
1:P:44:PHE:HB2	1:P:162:ALA:HB2	1.93	0.51
1:P:452:VAL:HG23	1:P:463:ILE:HG12	1.93	0.50
1:P:44:PHE:CD2	1:P:161:LYS:HB3	2.46	0.50
1:P:618:ASP:O	1:P:621:LYS:HB3	2.12	0.49
1:P:35:ILE:HG23	1:P:60:ARG:HH22	1.78	0.49
1:P:474:THR:HG23	1:P:493:THR:HB	1.94	0.48
1:P:494:LEU:N	1:P:494:LEU:HD23	2.28	0.48
1:P:666:LYS:HE2	1:P:670:ARG:HH12	1.77	0.48
1:P:181:ILE:HG22	1:P:188:MET:HE2	1.97	0.46
1:P:592:LYS:O	1:P:595:ASN:HB2	2.16	0.46
1:P:660:ILE:HA	1:P:663:LEU:HD12	1.98	0.45
1:P:791:ASN:HB2	1:P:811:VAL:HG21	1.99	0.45
1:P:694:LYS:HA	1:P:697:LYS:HD2	1.99	0.45
1:P:134:ASP:HB2	1:P:427:GLN:OE1	2.18	0.43
1:P:786:THR:O	1:P:789:VAL:HG12	2.17	0.43
1:P:620:GLU:HA	1:P:623:LEU:HD12	2.00	0.43
1:P:724:LEU:HD13	1:P:751:LEU:HD12	2.01	0.43
1:P:750:ALA:HB1	1:P:765:VAL:HG22	2.01	0.43
1:P:738:PRO:O	1:P:741:ILE:HG12	2.19	0.43
1:P:181:ILE:HG22	1:P:188:MET:CE	2.49	0.43
1:P:158:ASP:HA	1:P:161:LYS:HB2	2.00	0.42
1:P:73:TYR:OH	1:P:94:LYS:HD3	2.18	0.42
1:P:404:PHE:CD2	1:P:429:GLU:HB3	2.53	0.42
1:P:309:MET:HE1	1:P:381:TYR:OH	2.18	0.42
1:P:669:ASN:HD21	1:P:673:ARG:HD2	1.85	0.42
1:P:35:ILE:HG23	1:P:60:ARG:NH2	2.35	0.41
1:P:111:THR:HA	1:P:114:PHE:HD2	1.86	0.41
1:P:393:LEU:HB3	1:P:397:GLU:HB2	2.02	0.41
1:P:105:ILE:HD11	1:P:177:ILE:HD12	2.02	0.41
1:P:320:VAL:HG22	1:P:341:GLY:HA2	2.03	0.41
1:P:503:VAL:HG12	1:P:504:MET:N	2.36	0.41

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:P:706:PHE:HD1	1:P:710:LYS:HE3	1.85	0.41
1:P:657:LYS:HA	1:P:660:ILE:HD12	2.03	0.40

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	P	638/876 (73%)	593 (93%)	42 (7%)	3 (0%)	24 55

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	P	115	GLN
1	P	154	PRO
1	P	570	LEU

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	P	577/765 (75%)	517 (90%)	60 (10%)	7 25

All (60) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	P	26	PHE
1	P	30	LEU
1	P	35	ILE
1	P	78	ILE
1	P	82	THR
1	P	92	ILE
1	P	98	LEU
1	P	104	LEU
1	P	115	GLN
1	P	121	ARG
1	P	150	ILE
1	P	151	ILE
1	P	170	GLN
1	P	176	ILE
1	P	180	GLN
1	P	304	LEU
1	P	312	HIS
1	P	319	LYS
1	P	322	LEU
1	P	343	LYS
1	P	350	VAL
1	P	355	ASN
1	P	380	LEU
1	P	385	GLN
1	P	386	LEU
1	P	430	ILE
1	P	460	SER
1	P	467	ASN
1	P	471	LEU
1	P	472	SER
1	P	474	THR
1	P	493	THR
1	P	494	LEU
1	P	501	ASP
1	P	545	LEU
1	P	554	THR
1	P	558	ARG
1	P	568	LYS
1	P	613	LYS
1	P	619	LEU
1	P	648	GLU
1	P	657	LYS
1	P	679	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	P	688	LEU
1	P	699	LEU
1	P	706	PHE
1	P	722	GLU
1	P	723	THR
1	P	727	LEU
1	P	732	LYS
1	P	741	ILE
1	P	751	LEU
1	P	769	LYS
1	P	770	SER
1	P	779	VAL
1	P	800	LYS
1	P	807	ARG
1	P	814	ASP
1	P	822	GLN
1	P	823	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (12) such sidechains are listed below:

Mol	Chain	Res	Type
1	P	55	ASN
1	P	61	GLN
1	P	123	GLN
1	P	375	ASN
1	P	385	GLN
1	P	427	GLN
1	P	459	HIS
1	P	467	ASN
1	P	652	GLN
1	P	656	GLN
1	P	752	ASN
1	P	774	ASN

5.3.3 RNA ⓘ

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 oligosaccharides 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 [i](#)

6.1 Protein, DNA and RNA chains [i](#)

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	P	656/876 (74%)	-0.29	5 (0%)	82 68	80, 140, 217, 242	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	P	679	GLN	3.6
1	P	308	LEU	2.8
1	P	349	ASN	2.7
1	P	824	ALA	2.3
1	P	649	ALA	2.2

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 oligosaccharides 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.