



wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 15, 2025 – 11:13 am BST

PDB ID : 9EMH / pdb_00009emh
Title : KOD-H4 DNA polymerase mutant in a binary complex with DNA:DNA containing two HNA nucleotides
Authors : Gutfreund, C.; Betz, K.
Deposited on : 2024-03-08
Resolution : 2.70 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 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 Phenix2.0rc1
Mogul : 1.8.4, CSD as541be (2020)
Xtriage (Phenix) : 2.0rc1
EDS : 3.0
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4 : 9.0.006 (Gargrove)
Density-Fitness : 1.0.12
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.45.1

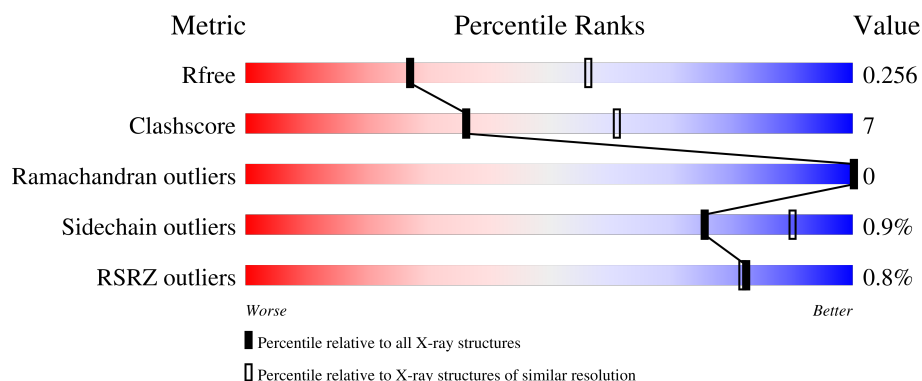
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION



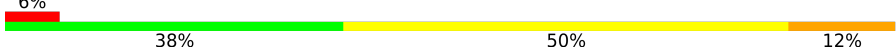
The reported resolution of this entry is 2.70 Å.

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}	164625	3333 (2.70-2.70)
Clashscore	180529	3684 (2.70-2.70)
Ramachandran outliers	177936	3633 (2.70-2.70)
Sidechain outliers	177891	3633 (2.70-2.70)
RSRZ outliers	164620	3333 (2.70-2.70)

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	A	774	
2	P	13	
3	T	16	

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 6861 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 DNA polymerase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	756	Total	C	N	O	S	0	1	0
			6206	3991	1055	1143	17			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	93	GLN	VAL	conflict	UNP D0VWU9
A	114	THR	ILE	conflict	UNP D0VWU9
A	141	ALA	ASP	conflict	UNP D0VWU9
A	143	ALA	GLU	conflict	UNP D0VWU9
A	147	HIS	GLU	conflict	UNP D0VWU9
A	383	LYS	SER	conflict	UNP D0VWU9
A	429	GLY	LYS	conflict	UNP D0VWU9
A	445	LEU	PHE	conflict	UNP D0VWU9
A	485	LEU	ALA	conflict	UNP D0VWU9
A	493	VAL	TYR	conflict	UNP D0VWU9
A	496	HIS	TYR	conflict	UNP D0VWU9
A	497	MET	TYR	conflict	UNP D0VWU9
A	499	PHE	TYR	conflict	UNP D0VWU9
A	500	GLU	ALA	conflict	UNP D0VWU9
A	501	ASN	ARG	conflict	UNP D0VWU9
A	521	LEU	ILE	conflict	UNP D0VWU9
A	584	LYS	GLU	conflict	UNP D0VWU9
A	664	LYS	GLU	conflict	UNP D0VWU9
A	726	ARG	LYS	conflict	UNP D0VWU9
A	735	LYS	ASN	conflict	UNP D0VWU9

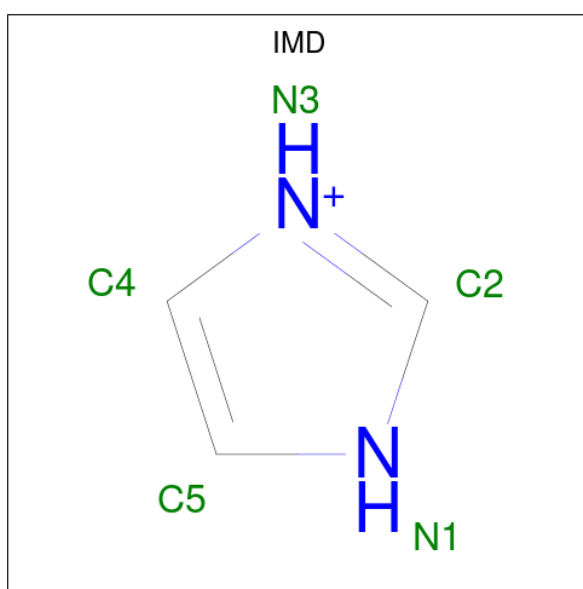
- Molecule 2 is a DNA chain called DNA (5'-D(*GP*AP*CP*CP*AP*CP*GP*GP*CP*CP*AP*CP*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	P	13	Total	C	N	O	P	0	0	0
			261	124	53	72	12			

- Molecule 3 is a DNA chain called DNA (5'-D(*AP*AP*CP*(6HT)P*(6HG)P*TP*GP*GP*CP*CP*GP*TP*GP*GP*TP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	T	16	Total	C	N	O	P	0	0	0
			329	158	60	96	15			

- Molecule 4 is IMIDAZOLE (CCD ID: IMD) (formula: C₃H₅N₂).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	N	0	0
			5	3	2		
4	A	1	Total	C	N	0	0
			5	3	2		
4	A	1	Total	C	N	0	0
			5	3	2		
4	A	1	Total	C	N	0	0
			5	3	2		

- Molecule 5 is MAGNESIUM ION (CCD ID: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	2	Total	Mg	0	0
			2	2		

- Molecule 6 is GLYCEROL (CCD ID: GOL) (formula: $C_3H_8O_3$).



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

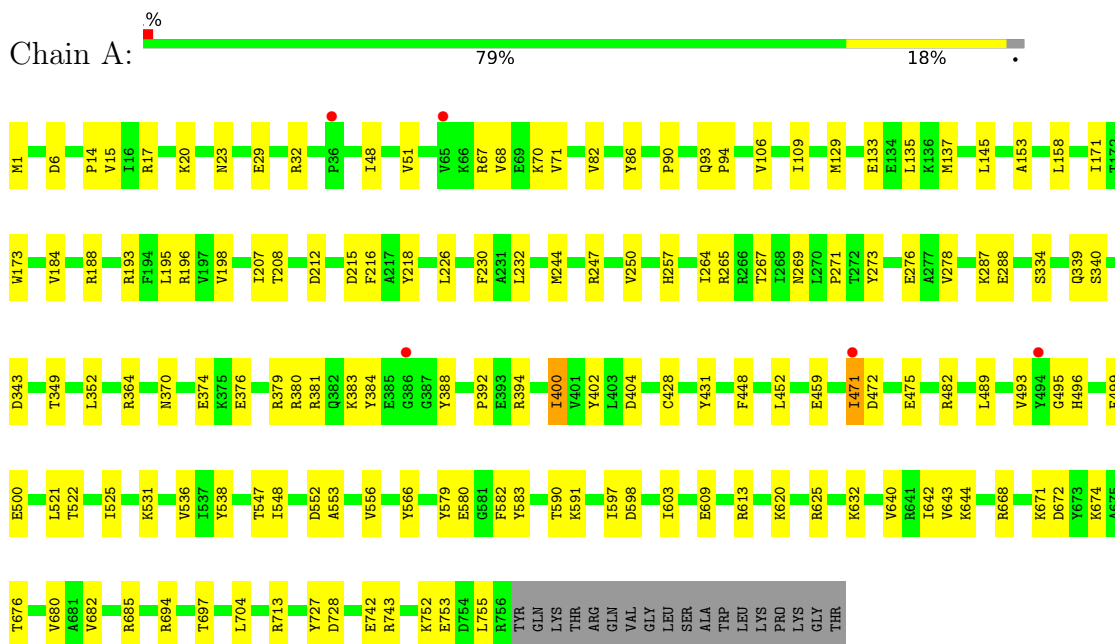
- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	33	Total	O	0	0
			33	33		
7	P	1	Total	O	0	0
			1	1		
7	T	3	Total	O	0	0
			3	3		

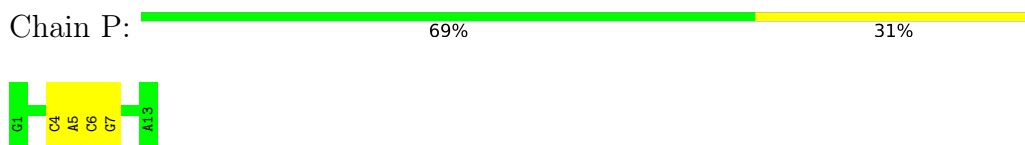
3 Residue-property plots [i](#)

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.

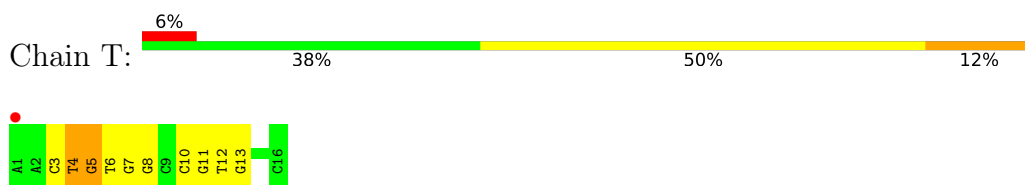
- Molecule 1: DNA polymerase



- Molecule 2: DNA (5'-D(*GP*AP*CP*CP*AP*CP*GP*GP*CP*CP*AP*CP*A)-3')



- Molecule 3: DNA (5'-D(*AP*AP*CP*(6HT)P*(6HG)P*TP*GP*GP*CP*CP*GP*TP*GP*GP*TP*C)-3')



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α , β , γ	111.48Å 151.62Å 67.58Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.03 – 2.70 46.03 – 2.70	Depositor EDS
% Data completeness (in resolution range)	99.3 (46.03-2.70) 99.8 (46.03-2.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.60 (at 2.69Å)	Xtriage
Refinement program	PHENIX 1.21.2_5419	Depositor
R, R_{free}	0.226 , 0.257 0.226 , 0.256	Depositor DCC
R_{free} test set	1188 reflections (2.20%)	wwPDB-VP
Wilson B-factor (Å ²)	66.0	Xtriage
Anisotropy	0.555	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 39.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	6861	wwPDB-VP
Average B, all atoms (Å ²)	79.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.08% 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: MG, IMD, 6HG, 6HT, GOL

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	0.11	0/6342	0.26	0/8552
2	P	0.18	0/293	0.32	0/449
3	T	0.17	0/318	0.35	0/487
All	All	0.12	0/6953	0.27	0/9488

There are no bond length outliers.

There are no bond angle outliers.

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	6206	0	6262	88	0
2	P	261	0	145	2	0
3	T	329	0	186	10	0
4	A	20	0	20	1	0
5	A	2	0	0	0	0
6	A	6	0	8	0	0
7	A	33	0	0	1	0
7	P	1	0	0	0	0
7	T	3	0	0	0	0
All	All	6861	0	6621	95	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 7.

The worst 5 of 95 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:743:ARG:NH2	3:T:10:DC:OP1	2.22	0.71
1:A:753:GLU:OE2	1:A:753:GLU:N	2.21	0.71
1:A:383:LYS:NZ	3:T:6:DT:OP2	2.24	0.70
1:A:364:ARG:HH21	1:A:452:LEU:HD23	1.64	0.61
1:A:247:ARG:HB3	1:A:694:ARG:HH22	1.65	0.61

There are no symmetry-related clashes.

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	755/774 (98%)	733 (97%)	22 (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	659/673 (98%)	653 (99%)	6 (1%)	75	90

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	288	GLU
1	A	400	ILE
1	A	471	ILE
1	A	106	VAL
1	A	70	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	23	ASN
1	A	91	GLN
1	A	257	HIS
1	A	304	ASN
1	A	501	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

2 non-standard protein/DNA/RNA residues 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	$\# Z > 2$
3	6HG	T	5	3,2	18,25,26	3.03	10 (55%)	19,36,39	2.21	7 (36%)
3	6HT	T	4	3	19,22,23	4.17	11 (57%)	24,31,34	3.04	11 (45%)

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
3	6HG	T	5	3,2	-	2/3/22/23	0/3/3/3
3	6HT	T	4	3	-	0/7/22/23	0/2/2/2

The worst 5 of 21 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	T	4	6HT	C6-C5	9.50	1.50	1.34
3	T	4	6HT	C2-N1	7.82	1.51	1.38
3	T	4	6HT	C2-N3	7.25	1.50	1.38
3	T	5	6HG	C2-N3	5.38	1.46	1.33
3	T	4	6HT	C6-N1	5.32	1.47	1.38

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	T	4	6HT	C6'-O4'-C4'	6.09	120.44	112.19
3	T	5	6HG	C3'-C2'-C1'	5.63	118.99	112.39
3	T	4	6HT	C3'-C2'-C1'	5.54	118.89	112.39
3	T	4	6HT	C5-C4-N3	5.36	119.89	115.31
3	T	4	6HT	C4-N3-C2	-5.35	120.43	127.35

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	T	5	6HG	O4'-C4'-C5'-O5'
3	T	5	6HG	C3'-C4'-C5'-O5'

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	T	5	6HG	2	0
3	T	4	6HT	2	0

5.5 Carbohydrates

There are no oligosaccharides in this entry.

5.6 Ligand geometry

Of 7 ligands modelled in this entry, 2 are monoatomic - leaving 5 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	$\# Z > 2$
4	IMD	A	804	-	3,5,5	0.16	0	4,5,5	1.00	0
4	IMD	A	807	-	3,5,5	0.16	0	4,5,5	0.99	0
4	IMD	A	805	-	3,5,5	0.15	0	4,5,5	0.98	0
4	IMD	A	801	-	3,5,5	0.16	0	4,5,5	1.01	0
6	GOL	A	806	-	5,5,5	0.34	0	5,5,5	0.37	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
4	IMD	A	807	-	-	-	0/1/1/1
4	IMD	A	801	-	-	-	0/1/1/1
4	IMD	A	805	-	-	-	0/1/1/1
6	GOL	A	806	-	-	0/4/4/4	-
4	IMD	A	804	-	-	-	0/1/1/1

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.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	805	IMD	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 [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	A	756/774 (97%)	0.05	5 (0%) 84 83	29, 76, 108, 132	1 (0%)
2	P	13/13 (100%)	-0.19	0 100 100	79, 94, 117, 123	0
3	T	14/16 (87%)	0.02	1 (7%) 23 21	77, 91, 106, 130	0
All	All	783/803 (97%)	0.04	6 (0%) 82 82	29, 77, 108, 132	1 (0%)

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	471	ILE	2.9
1	A	386	GLY	2.9
3	T	1	DA	2.7
1	A	36	PRO	2.4
1	A	494	TYR	2.2

6.2 Non-standard residues in protein, DNA, RNA chains [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
3	6HT	T	4	21/22	0.89	0.10	88,98,107,110	0
3	6HG	T	5	23/24	0.90	0.08	77,91,96,110	0

6.3 Carbohydrates [i](#)

There are no oligosaccharides 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
4	IMD	A	805	5/5	0.78	0.25	87,89,92,93	0
4	IMD	A	807	5/5	0.80	0.20	84,86,93,100	0
4	IMD	A	804	5/5	0.81	0.20	77,83,87,94	0
5	MG	A	803	1/1	0.85	0.15	95,95,95,95	0
6	GOL	A	806	6/6	0.88	0.12	84,87,88,92	0
4	IMD	A	801	5/5	0.89	0.15	89,91,94,94	0
5	MG	A	802	1/1	0.96	0.07	70,70,70,70	0

6.5 Other polymers [i](#)

There are no such residues in this entry.