wwPDB X-ray Structure Validation Summary Report (i)

Oct 16, 2021 – 10:08 PM EDT

PDB ID	:	1NOP
Title	:	Crystal structure of human tyrosyl-DNA phosphodiesterase (Tdp1) in complex with vanadate, DNA and a human topoisomerase I-derived peptide
Authors	:	Davies, D.R.; Interthal, H.; Champoux, J.J.; Hol, W.G.J.
Deposited on	:	2003-01-16
Resolution	:	2.30 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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

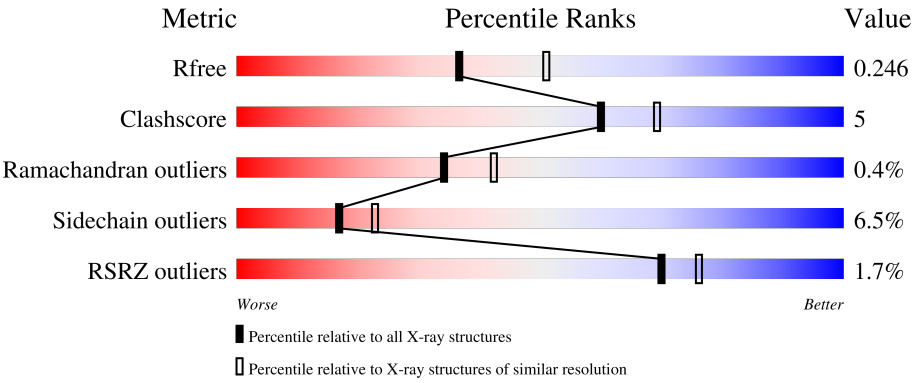
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 2.30 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}	130704	5042 (2.30-2.30)
Clashscore	141614	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)
RSRZ outliers	127900	4938 (2.30-2.30)

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	D	6	17% 33% 17% 50%
1	F	6	17% 83%
2	A	485	2% 74% 13% • 12%
2	B	485	% 74% 13% • 11%
3	C	8	50% 12% 38%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 7072 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 DNA chain called 5'-D(*AP*GP*AP*GP*TP*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	D	3	Total	C	N	O	P	0	0	0
			62	30	9	20	3			
1	F	1	Total	C	N	O	P	0	0	0
			20	10	2	7	1			

- Molecule 2 is a protein called tyrosyl-DNA phosphodiesterase 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	A	425	Total	C	N	O	S	0	0	0
			3397	2210	574	602	11			
2	B	433	Total	C	N	O	S	0	0	0
			3446	2239	583	613	11			

There are 56 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	124	MET	-	cloning artifact	GB 20127586
A	125	GLY	-	cloning artifact	GB 20127586
A	126	SER	-	cloning artifact	GB 20127586
A	127	SER	-	cloning artifact	GB 20127586
A	128	HIS	-	cloning artifact	GB 20127586
A	129	HIS	-	cloning artifact	GB 20127586
A	130	HIS	-	cloning artifact	GB 20127586
A	131	HIS	-	cloning artifact	GB 20127586
A	132	HIS	-	cloning artifact	GB 20127586
A	133	HIS	-	cloning artifact	GB 20127586
A	134	SER	-	cloning artifact	GB 20127586
A	135	SER	-	cloning artifact	GB 20127586
A	136	GLY	-	cloning artifact	GB 20127586
A	137	LEU	-	cloning artifact	GB 20127586
A	138	VAL	-	cloning artifact	GB 20127586
A	139	PRO	-	cloning artifact	GB 20127586

Continued on next page...

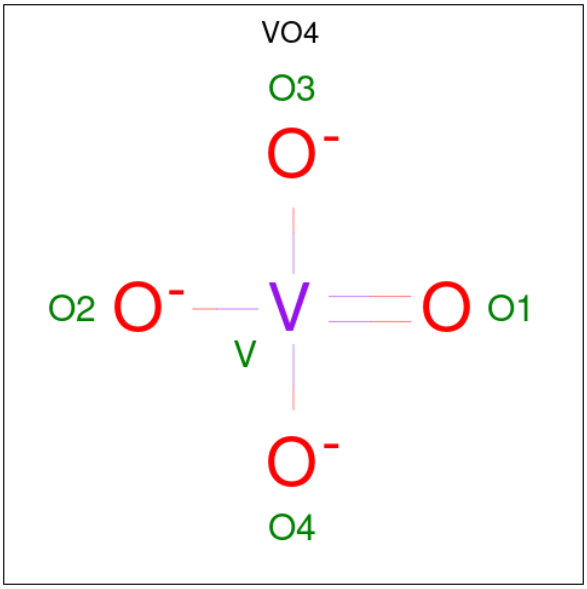
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	140	ARG	-	cloning artifact	GB 20127586
A	141	GLY	-	cloning artifact	GB 20127586
A	142	SER	-	cloning artifact	GB 20127586
A	143	HIS	-	cloning artifact	GB 20127586
A	144	MET	-	cloning artifact	GB 20127586
A	145	LEU	-	cloning artifact	GB 20127586
A	146	GLU	-	cloning artifact	GB 20127586
A	147	ASP	-	cloning artifact	GB 20127586
A	148	PRO	-	cloning artifact	GB 20127586
A	322	ASN	ASP	engineered mutation	GB 20127586
A	328	THR	MET	engineered mutation	GB 20127586
A	548	LEU	PHE	engineered mutation	GB 20127586
B	124	MET	-	cloning artifact	GB 20127586
B	125	GLY	-	cloning artifact	GB 20127586
B	126	SER	-	cloning artifact	GB 20127586
B	127	SER	-	cloning artifact	GB 20127586
B	128	HIS	-	cloning artifact	GB 20127586
B	129	HIS	-	cloning artifact	GB 20127586
B	130	HIS	-	cloning artifact	GB 20127586
B	131	HIS	-	cloning artifact	GB 20127586
B	132	HIS	-	cloning artifact	GB 20127586
B	133	HIS	-	cloning artifact	GB 20127586
B	134	SER	-	cloning artifact	GB 20127586
B	135	SER	-	cloning artifact	GB 20127586
B	136	GLY	-	cloning artifact	GB 20127586
B	137	LEU	-	cloning artifact	GB 20127586
B	138	VAL	-	cloning artifact	GB 20127586
B	139	PRO	-	cloning artifact	GB 20127586
B	140	ARG	-	cloning artifact	GB 20127586
B	141	GLY	-	cloning artifact	GB 20127586
B	142	SER	-	cloning artifact	GB 20127586
B	143	HIS	-	cloning artifact	GB 20127586
B	144	MET	-	cloning artifact	GB 20127586
B	145	LEU	-	cloning artifact	GB 20127586
B	146	GLU	-	cloning artifact	GB 20127586
B	147	ASP	-	cloning artifact	GB 20127586
B	148	PRO	-	cloning artifact	GB 20127586
B	322	ASN	ASP	engineered mutation	GB 20127586
B	328	THR	MET	engineered mutation	GB 20127586
B	548	LEU	PHE	engineered mutation	GB 20127586

- Molecule 3 is a protein called topoisomerase I-derived peptide.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
3	C	5	Total C N O 45 31 7 7	0	0	0

- Molecule 4 is VANADATE ION (three-letter code: VO4) (formula: O_4V).



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

- Molecule 5 is water.

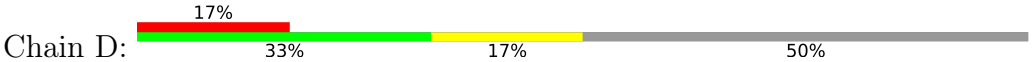
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	D	1	Total O 1 1	0	0
5	A	44	Total O 44 44	0	0
5	B	50	Total O 50 50	0	0
5	C	1	Total O 1 1	0	0

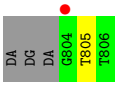
3 Residue-property plots (i)

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: 5'-D(*AP*GP*AP*GP*TP*T)-3'

Chain D:



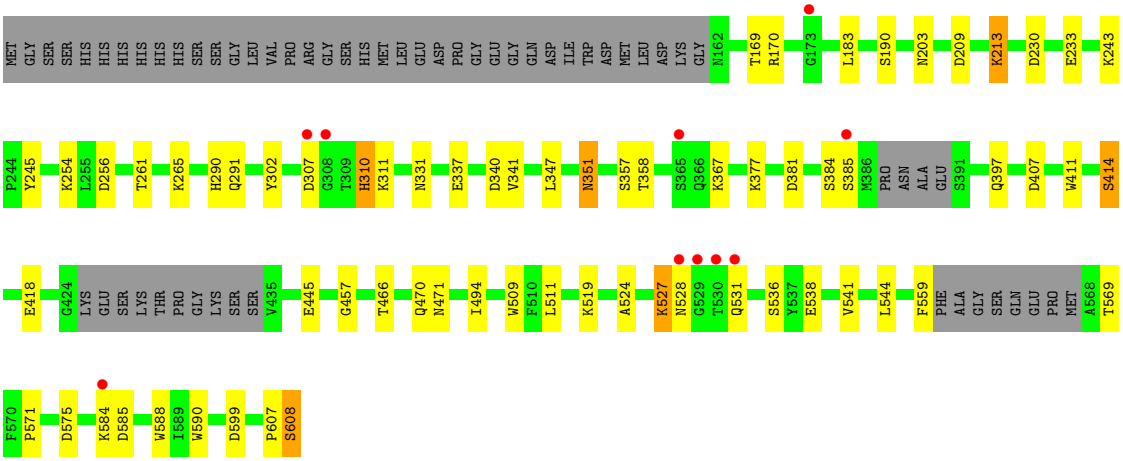


• Molecule 1: 5'-D(*AP*GP*AP*GP*TP*T)-3'

Chain F:

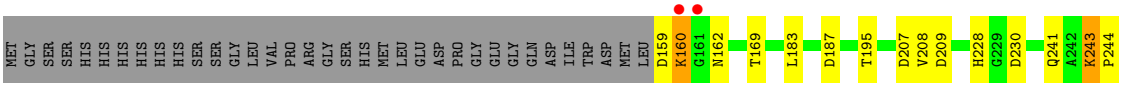
• Molecule 2: tyrosyl-DNA phosphodiesterase 1

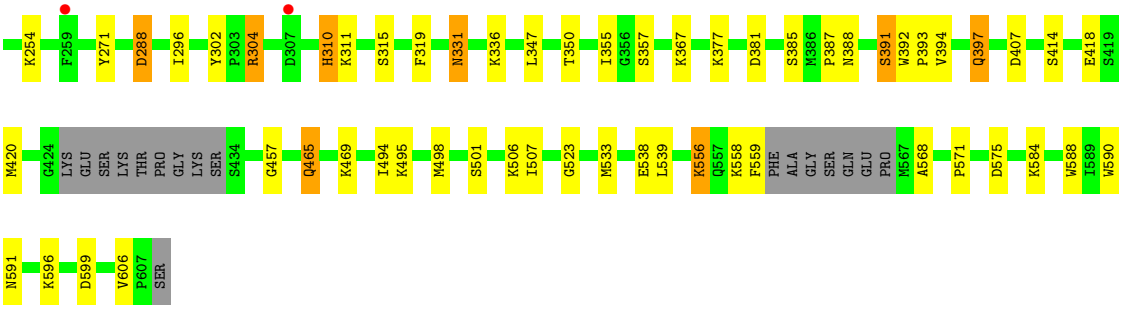
Chain A:



• Molecule 2: tyrosyl-DNA phosphodiesterase 1

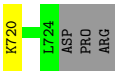
Chain B:





• Molecule 3: topoisomerase I-derived peptide

Chain C:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	49.80Å 104.72Å 193.92Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	95.35 – 2.30 48.24 – 2.30	Depositor EDS
% Data completeness (in resolution range)	93.7 (95.35-2.30) 93.7 (48.24-2.30)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.09	Depositor
$< I/\sigma(I) >$ [1]	2.73 (at 2.29Å)	Xtriage
Refinement program	REFMAC 5.1.24	Depositor
R, R_{free}	0.206 , 0.252 0.205 , 0.246	Depositor DCC
R_{free} test set	2169 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	30.2	Xtriage
Anisotropy	0.140	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 35.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	7072	wwPDB-VP
Average B, all atoms (Å^2)	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.07% 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: VO4

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	D	1.44	0/68	2.04	3/103 (2.9%)
1	F	1.18	0/21	1.68	0/30
2	A	0.59	0/3508	0.77	7/4770 (0.1%)
2	B	0.63	0/3559	0.77	6/4841 (0.1%)
3	C	0.64	0/45	0.80	0/59
All	All	0.63	0/7201	0.80	16/9803 (0.2%)

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
2	A	0	1

There are no bond length outliers.

The worst 5 of 16 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	340	ASP	CB-CG-OD2	10.27	127.54	118.30
2	B	599	ASP	CB-CG-OD2	8.16	125.65	118.30
2	A	209	ASP	CB-CG-OD2	7.59	125.14	118.30
2	B	209	ASP	CB-CG-OD2	7.58	125.13	118.30
2	B	407	ASP	CB-CG-OD2	7.34	124.90	118.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	A	527	LYS	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	D	62	0	35	0	0
1	F	20	0	12	0	0
2	A	3397	0	3336	27	0
2	B	3446	0	3372	36	0
3	C	45	0	48	1	0
4	A	3	0	0	0	0
4	B	3	0	0	1	0
5	A	44	0	0	1	0
5	B	50	0	0	0	0
5	C	1	0	0	0	0
5	D	1	0	0	0	0
All	All	7072	0	6803	63	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.

The worst 5 of 63 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:B:304:ARG:HH21	2:B:304:ARG:HG2	1.48	0.78
2:B:228:HIS:HD2	2:B:230:ASP:H	1.33	0.75
2:B:498:MET:CE	2:B:507:ILE:HG21	2.18	0.73
2:A:290:HIS:CD2	2:A:291:GLN:HG3	2.25	0.72
2:A:357:SER:HB2	2:A:538:GLU:HB2	1.73	0.69

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
2	A	417/485 (86%)	400 (96%)	14 (3%)	3 (1%)	22 26
2	B	427/485 (88%)	414 (97%)	13 (3%)	0	100 100
3	C	3/8 (38%)	2 (67%)	1 (33%)	0	100 100
All	All	847/978 (87%)	816 (96%)	28 (3%)	3 (0%)	34 42

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	A	307	ASP
2	A	414	SER
2	A	411	TRP

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
2	A	367/421 (87%)	344 (94%)	23 (6%)	18 24
2	B	370/421 (88%)	345 (93%)	25 (7%)	16 21
3	C	5/8 (62%)	5 (100%)	0	100 100
All	All	742/850 (87%)	694 (94%)	48 (6%)	17 23

5 of 48 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	288	ASP
2	B	367	LYS
2	B	304	ARG
2	B	315	SER
2	B	391	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	465	GLN
2	B	397	GLN
2	B	165	GLN
2	B	331	ASN
2	A	531	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)

2 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	VO4	B	699	2,1	0,2,4	-	-	-		
4	VO4	A	699	2,3,1	0,2,4	-	-	-		

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.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	B	699	VO4	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

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	D	3/6 (50%)	0.70	1 (33%) 0 0	28, 28, 46, 59	0
1	F	1/6 (16%)	0.92	0 100 100	59, 59, 59, 59	0
2	A	425/485 (87%)	-0.23	10 (2%) 59 66	14, 30, 48, 66	0
2	B	433/485 (89%)	-0.25	4 (0%) 84 88	11, 25, 43, 71	0
3	C	5/8 (62%)	-0.46	0 100 100	26, 30, 32, 33	0
All	All	867/990 (87%)	-0.24	15 (1%) 70 76	11, 27, 47, 71	0

The worst 5 of 15 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	A	528	ASN	6.1
2	A	365	SER	4.6
1	D	804	DG	3.6
2	B	307	ASP	3.5
2	A	531	GLN	3.5

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, 95^{th} 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	VO4	B	699	3/5	0.99	0.12	33,33,33,33	0
4	VO4	A	699	3/5	1.00	0.12	26,26,27,27	0

6.5 Other polymers (i)

There are no such residues in this entry.