



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 10, 2018 – 06:19 am GMT

PDB ID : 2PYL
Title : Phi29 DNA polymerase complexed with primer-template DNA and incoming nucleotide substrates (ternary complex)
Authors : Berman, A.J.; Kamtekar, S.; Goodman, J.L.; Lazaro, J.M.; de Vega, M.; Blanco, L.; Salas, M.; Steitz, T.A.
Deposited on : 2007-05-16
Resolution : 2.20 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.7.3 (157068), CSD as539be (2018)
Xtriage (Phenix) : 1.13
EDS : trunk30967
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac : 5.8.0158
CCP4 : 7.0 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : trunk30967

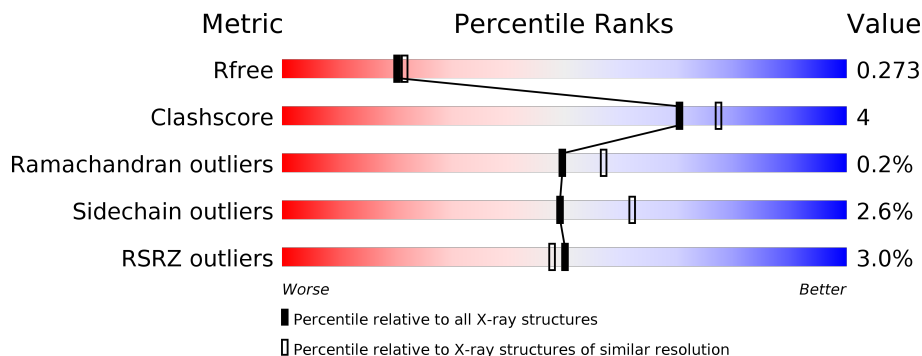
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.20 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}	111664	4343 (2.20-2.20)
Clashscore	122126	5027 (2.20-2.20)
Ramachandran outliers	120053	4952 (2.20-2.20)
Sidechain outliers	120020	4953 (2.20-2.20)
RSRZ outliers	108989	4245 (2.20-2.20)

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 on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. 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	X	11	<div> <div>27%</div> <div>18%</div> <div>45%</div> <div>36%</div> </div>
2	Y	14	<div> <div>21%</div> <div>64%</div> <div>29%</div> <div>7%</div> </div>
3	A	575	<div> <div>2%</div> <div>88%</div> <div>10%</div> <div>..</div> </div>

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 5597 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(GACTGCTTAC(2DA))-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	X	11	Total	C	N	O	P	0	0	0
			220	107	40	63	10			

- Molecule 2 is a DNA chain called 5'-d(CTGACGAATGTACA)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	Y	14	Total	C	N	O	P	0	0	0
			285	137	55	80	13			

- Molecule 3 is a protein called DNA polymerase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	571	Total	C	N	O	S	4	2	0
			4680	3048	755	856	21			

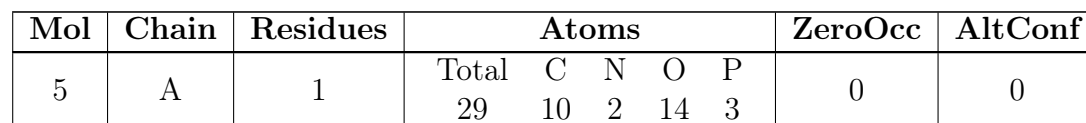
There are 2 discrepancies between the modelled and reference sequences:


Chain	Residue	Modelled	Actual	Comment	Reference
A	12	ALA	ASP	ENGINEERED	UNP P03680
A	66	ALA	ASP	ENGINEERED	UNP P03680

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

- Molecule 5 is THYMIDINE-5'-TRIPHOSPHATE (three-letter code: TTP) (formula: C₁₀H₁₇N₂O₁₄P₃).



- 
- The diagram shows the chemical structure of EDO (Ethane-1,2-diol). It features a zigzag line representing the carbon chain, with two vertices labeled C1 and C2 in green. Each vertex is connected to a hydroxyl group (OH) in red. The hydroxyl group at C1 is labeled O1 in green, and the hydroxyl group at C2 is labeled O2 in green. The entire structure is titled "EDO" in black at the top center.

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



Continued from previous page...

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

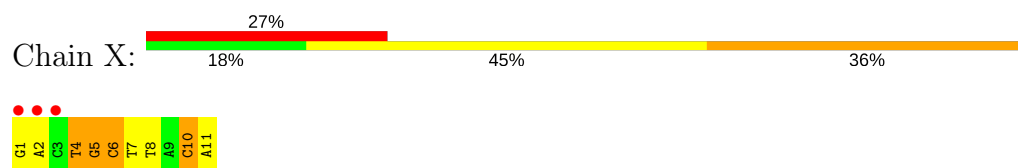
- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	X	11	Total	O	0	0
			11	11		
7	Y	23	Total	O	0	0
			23	23		
7	A	323	Total	O	0	0
			323	323		

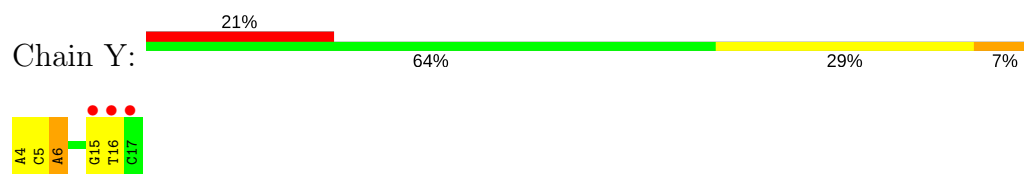
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA 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.

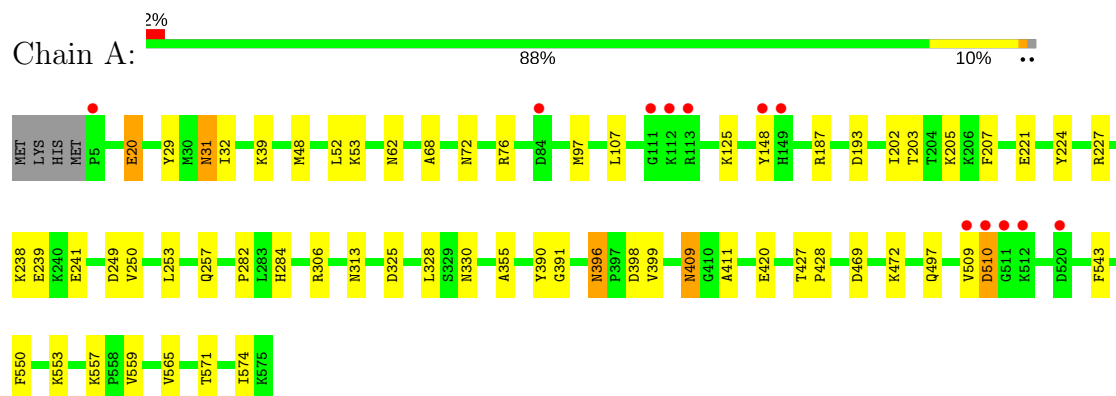
- Molecule 1: 5'-d(GACTGCTTAC(2DA))-3'



- Molecule 2: 5'-d(CTGACGAATGTACA)-3'



- Molecule 3: DNA polymerase



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	58.92Å 78.20Å 157.81Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.42 – 2.20 40.42 – 2.20	Depositor EDS
% Data completeness (in resolution range)	96.9 (40.42-2.20) 96.9 (40.42-2.20)	Depositor EDS
R_{merge}	0.17	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.83 (at 2.20Å)	Xtriage
Refinement program	REFMAC 5.2	Depositor
R, R_{free}	0.191 , 0.253 0.217 , 0.273	Depositor DCC
R_{free} test set	3649 reflections (9.95%)	wwPDB-VP
Wilson B-factor (Å ²)	30.3	Xtriage
Anisotropy	0.294	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 34.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	5597	wwPDB-VP
Average B, all atoms (Å ²)	17.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.40% 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, TTP, 2DA, EDO

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	X	1.29	1/223 (0.4%)	1.87	10/342 (2.9%)
2	Y	0.86	0/320	1.50	6/492 (1.2%)
3	A	0.52	1/4803 (0.0%)	0.60	2/6480 (0.0%)
All	All	0.60	2/5346 (0.0%)	0.79	18/7314 (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
3	A	0	1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	306	ARG	CD-NE	-9.78	1.29	1.46
1	X	10	DC	N3-C4	-5.12	1.30	1.33

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	X	7	DT	O4'-C4'-C3'	-9.09	100.55	106.00
1	X	5	DG	P-O3'-C3'	8.07	129.38	119.70
1	X	4	DT	P-O3'-C3'	7.96	129.25	119.70
2	Y	6	DA	O5'-P-OP1	-7.65	98.81	105.70
1	X	2	DA	P-O3'-C3'	7.03	128.14	119.70

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	A	20	GLU	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	X	220	0	126	3	0
2	Y	285	0	159	1	0
3	A	4680	0	4686	37	0
4	A	2	0	0	0	0
5	A	29	0	13	0	0
6	A	24	0	36	2	0
7	A	323	0	0	0	0
7	X	11	0	0	0	0
7	Y	23	0	0	0	0
All	All	5597	0	5020	40	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 4.

The worst 5 of 40 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)
3:A:509:VAL:HG13	3:A:510:ASP:H	1.17	1.09
3:A:509:VAL:CG1	3:A:510:ASP:H	1.84	0.91
1:X:10:DC:H2''	1:X:11:2DA:H5''	1.59	0.84
3:A:72:ASN:HD22	3:A:76:ARG:HH12	1.29	0.79
3:A:509:VAL:HG13	3:A:510:ASP:N	1.96	0.77

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
3	A	571/575 (99%)	552 (97%)	18 (3%)	1 (0%)	49 57

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	62	ASN

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
3	A	504/506 (100%)	491 (97%)	13 (3%)	49 62

5 of 13 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	241	GLU
3	A	249	ASP
3	A	409	ASN
3	A	148	TYR
3	A	396	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
3	A	380	GLN
3	A	497	GLN
3	A	396	ASN
3	A	62	ASN
3	A	409	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

1 non-standard protein/DNA/RNA residue is 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$
1	2DA	X	11	1,2	17,22,23	1.73	4 (23%)	13,31,34	2.77	7 (53%)

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
1	2DA	X	11	1,2	-	0/3/18/19	0/3/3/3

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	X	11	2DA	O5'-C5'	-4.26	1.38	1.44
1	X	11	2DA	C8-N9	-3.01	1.33	1.36
1	X	11	2DA	C4-N3	-2.13	1.32	1.35
1	X	11	2DA	O4'-C4'	-2.06	1.40	1.44

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	X	11	2DA	N3-C2-N1	-6.17	123.58	128.86
1	X	11	2DA	C4-C5-N7	-3.15	106.36	109.41
1	X	11	2DA	C2'-C1'-N9	-2.93	106.96	112.48
1	X	11	2DA	O4'-C4'-C5'	-2.48	105.57	109.59
1	X	11	2DA	O4'-C4'-C3'	2.38	108.77	104.80

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
1	X	11	2DA	1	0

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 9 ligands modelled in this entry, 2 are monoatomic - leaving 7 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	$\# Z > 2$
6	EDO	A	1001	-	3,3,3	0.58	0	2,2,2	0.19	0
6	EDO	A	1002	-	3,3,3	0.35	0	2,2,2	0.52	0
6	EDO	A	1003	-	3,3,3	0.53	0	2,2,2	0.29	0
6	EDO	A	1004	-	3,3,3	0.74	0	2,2,2	0.12	0
6	EDO	A	1005	-	3,3,3	0.50	0	2,2,2	0.30	0
6	EDO	A	1006	-	3,3,3	0.43	0	2,2,2	0.41	0
5	TTP	A	3204	4	22,30,30	0.72	0	27,47,47	1.87	5 (18%)

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
6	EDO	A	1001	-	-	0/1/1/1	0/0/0/0
6	EDO	A	1002	-	-	0/1/1/1	0/0/0/0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
6	EDO	A	1003	-	-	0/1/1/1	0/0/0/0
6	EDO	A	1004	-	-	0/1/1/1	0/0/0/0
6	EDO	A	1005	-	-	0/1/1/1	0/0/0/0
6	EDO	A	1006	-	-	0/1/1/1	0/0/0/0
5	TTP	A	3204	4	-	0/18/34/34	0/2/2/2

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	3204	TTP	C2'-C1'-N1	-2.94	107.41	114.27
5	A	3204	TTP	PB-O3B-PG	-2.79	123.24	132.63
5	A	3204	TTP	PB-O3A-PA	-2.25	125.08	132.63
5	A	3204	TTP	O4'-C1'-N1	2.19	111.47	107.78
5	A	3204	TTP	C4-N3-C2	7.78	121.77	115.14

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	A	1002	EDO	2	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	X	10/11 (90%)	0.90	3 (30%) 0 0	15, 17, 47, 51	0
2	Y	14/14 (100%)	0.61	3 (21%) 1 0	13, 17, 35, 40	0
3	A	571/575 (99%)	-0.16	12 (2%) 63 61	9, 16, 25, 40	1 (0%)
All	All	595/600 (99%)	-0.12	18 (3%) 50 48	9, 16, 26, 51	1 (0%)

The worst 5 of 18 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	Y	17	DC	5.5
3	A	5	PRO	4.4
3	A	510	ASP	3.9
3	A	148	TYR	3.8
3	A	112	LYS	3.6

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
1	2DA	X	11	20/21	0.96	0.15	13,17,19,20	0

6.3 Carbohydrates [i](#)

There are no carbohydrates 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(\AA^2)	Q<0.9
6	EDO	A	1001	4/4	0.67	0.21	34,36,37,38	0
6	EDO	A	1004	4/4	0.69	0.23	8,12,14,14	0
4	MG	A	3203	1/1	0.79	0.16	19,19,19,19	0
6	EDO	A	1003	4/4	0.80	0.15	38,39,39,39	0
6	EDO	A	1002	4/4	0.84	0.18	19,19,21,23	0
6	EDO	A	1006	4/4	0.89	0.19	22,24,25,27	0
6	EDO	A	1005	4/4	0.89	0.14	33,34,34,34	0
4	MG	A	3202	1/1	0.96	0.08	22,22,22,22	0
5	TTP	A	3204	29/29	0.96	0.11	2,9,18,20	0

6.5 Other polymers [i](#)

There are no such residues in this entry.