



wwPDB X-ray Structure Validation Summary Report ⓘ

May 15, 2020 – 01:56 am BST

PDB ID : 3SQ7
Title : Crystal Structure Analysis of the Yeast Tyrosyl-DNA Phosphodiesterase H432N_Glu Mutant
Authors : Gajewski, S.; White, S.W.
Deposited on : 2011-07-05
Resolution : 2.00 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	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
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)	:	2.11

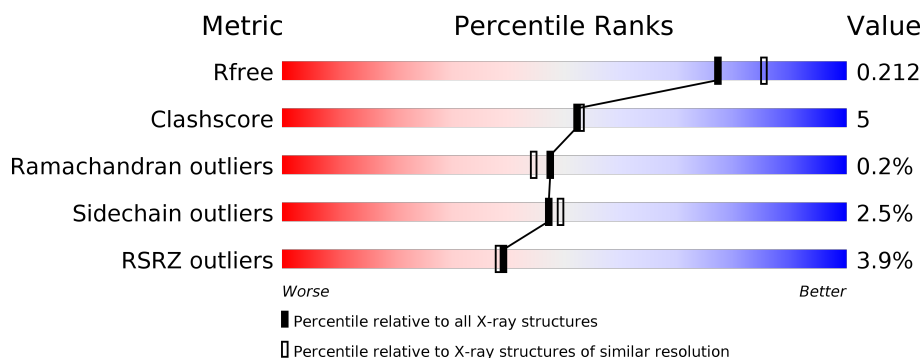
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.00 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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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 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	470	<div> <div>3%</div> <div> <div></div> <div>80%</div> <div>11%</div> <div>8%</div> </div> </div>
1	B	470	<div> <div>5%</div> <div> <div></div> <div>78%</div> <div>12%</div> <div>9%</div> </div> </div>
1	C	470	<div> <div>3%</div> <div> <div></div> <div>80%</div> <div>10%</div> <div>9%</div> </div> </div>
1	D	470	<div> <div>4%</div> <div> <div></div> <div>78%</div> <div>11%</div> <div>10%</div> </div> </div>

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 14719 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 Tyrosyl-DNA phosphodiesterase 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	432	Total	C	N	O	S	0	2	0
			3502	2268	573	639	22			
1	B	426	Total	C	N	O	S	0	3	0
			3469	2248	569	629	23			
1	C	430	Total	C	N	O	S	0	3	0
			3503	2270	575	636	22			
1	D	425	Total	C	N	O	S	0	3	0
			3466	2248	570	625	23			

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	78	MET	-	INITIATING METHIONINE	UNP P38319
A	432	ASN	HIS	ENGINEERED MUTATION	UNP P38319
A	540	LEU	-	EXPRESSION TAG	UNP P38319
A	541	GLU	-	EXPRESSION TAG	UNP P38319
A	542	HIS	-	EXPRESSION TAG	UNP P38319
A	543	HIS	-	EXPRESSION TAG	UNP P38319
A	544	HIS	-	EXPRESSION TAG	UNP P38319
A	545	HIS	-	EXPRESSION TAG	UNP P38319
A	546	HIS	-	EXPRESSION TAG	UNP P38319
A	547	HIS	-	EXPRESSION TAG	UNP P38319
B	78	MET	-	INITIATING METHIONINE	UNP P38319
B	432	ASN	HIS	ENGINEERED MUTATION	UNP P38319
B	540	LEU	-	EXPRESSION TAG	UNP P38319
B	541	GLU	-	EXPRESSION TAG	UNP P38319
B	542	HIS	-	EXPRESSION TAG	UNP P38319
B	543	HIS	-	EXPRESSION TAG	UNP P38319
B	544	HIS	-	EXPRESSION TAG	UNP P38319
B	545	HIS	-	EXPRESSION TAG	UNP P38319
B	546	HIS	-	EXPRESSION TAG	UNP P38319
B	547	HIS	-	EXPRESSION TAG	UNP P38319
C	78	MET	-	INITIATING METHIONINE	UNP P38319

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	432	ASN	HIS	ENGINEERED MUTATION	UNP P38319
C	540	LEU	-	EXPRESSION TAG	UNP P38319
C	541	GLU	-	EXPRESSION TAG	UNP P38319
C	542	HIS	-	EXPRESSION TAG	UNP P38319
C	543	HIS	-	EXPRESSION TAG	UNP P38319
C	544	HIS	-	EXPRESSION TAG	UNP P38319
C	545	HIS	-	EXPRESSION TAG	UNP P38319
C	546	HIS	-	EXPRESSION TAG	UNP P38319
C	547	HIS	-	EXPRESSION TAG	UNP P38319
D	78	MET	-	INITIATING METHIONINE	UNP P38319
D	432	ASN	HIS	ENGINEERED MUTATION	UNP P38319
D	540	LEU	-	EXPRESSION TAG	UNP P38319
D	541	GLU	-	EXPRESSION TAG	UNP P38319
D	542	HIS	-	EXPRESSION TAG	UNP P38319
D	543	HIS	-	EXPRESSION TAG	UNP P38319
D	544	HIS	-	EXPRESSION TAG	UNP P38319
D	545	HIS	-	EXPRESSION TAG	UNP P38319
D	546	HIS	-	EXPRESSION TAG	UNP P38319
D	547	HIS	-	EXPRESSION TAG	UNP P38319

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	O	S	0	0
			5	4	1		
2	A	1	Total	O	S	0	0
			5	4	1		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	O	S	0	0
			5	4	1		
2	A	1	Total	O	S	0	0
			5	4	1		
2	B	1	Total	O	S	0	0
			5	4	1		
2	D	1	Total	O	S	0	0
			5	4	1		
2	D	1	Total	O	S	0	0
			5	4	1		

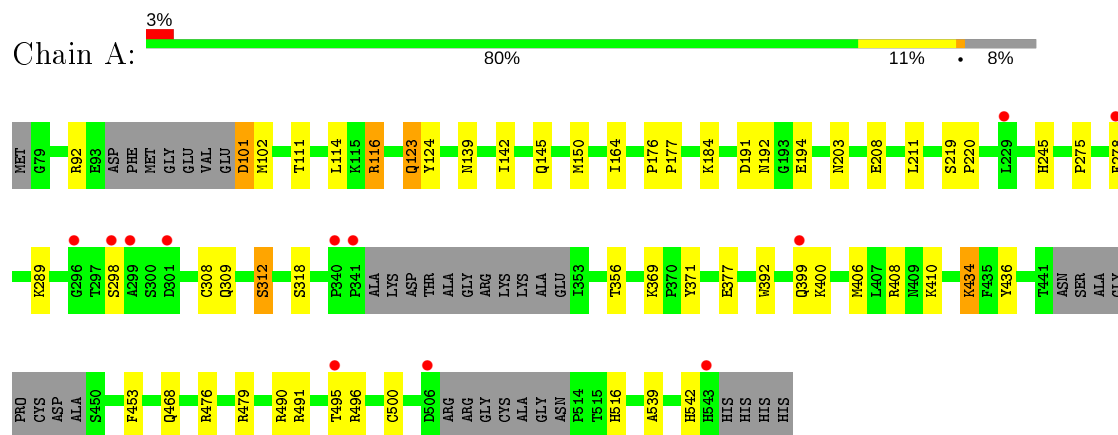
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	237	Total	O	0	0
			237	237		
3	B	151	Total	O	0	0
			151	151		
3	C	199	Total	O	0	0
			199	199		
3	D	157	Total	O	0	0
			157	157		

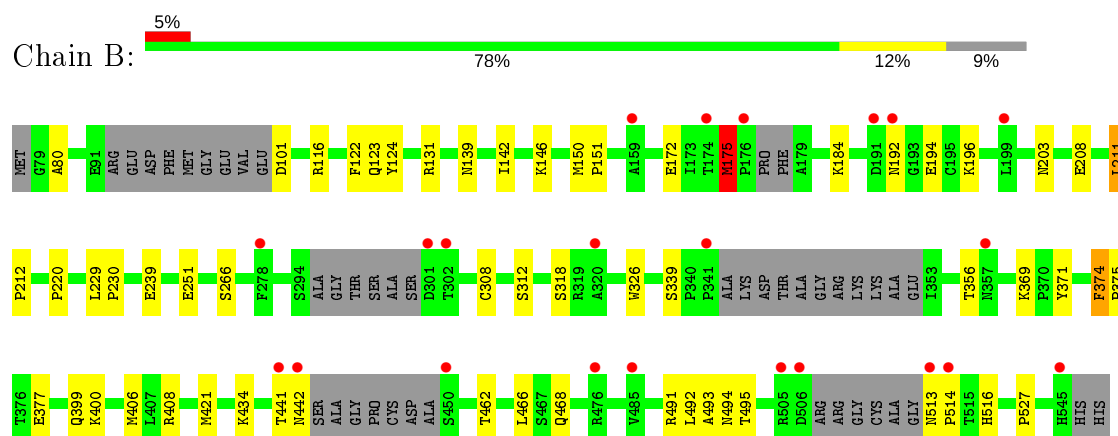
3 Residue-property plots [i](#)

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.

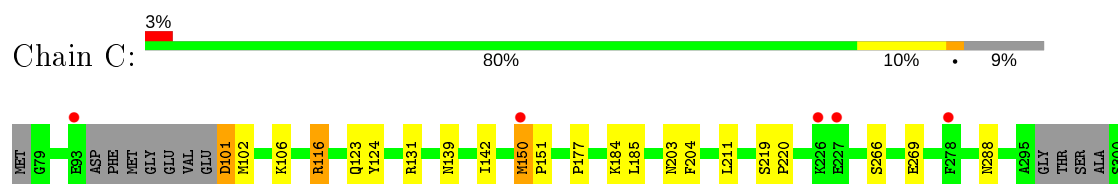
• Molecule 1: Tyrosyl-DNA phosphodiesterase 1

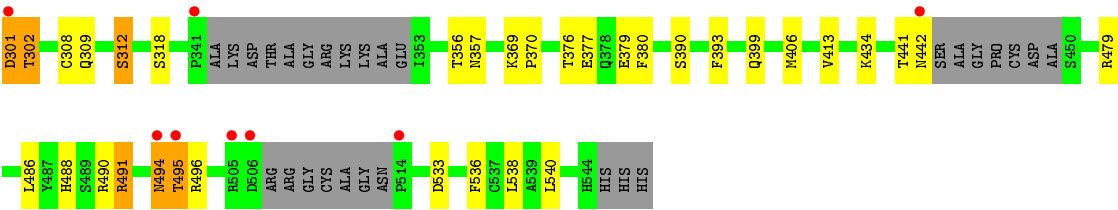


• Molecule 1: Tyrosyl-DNA phosphodiesterase 1

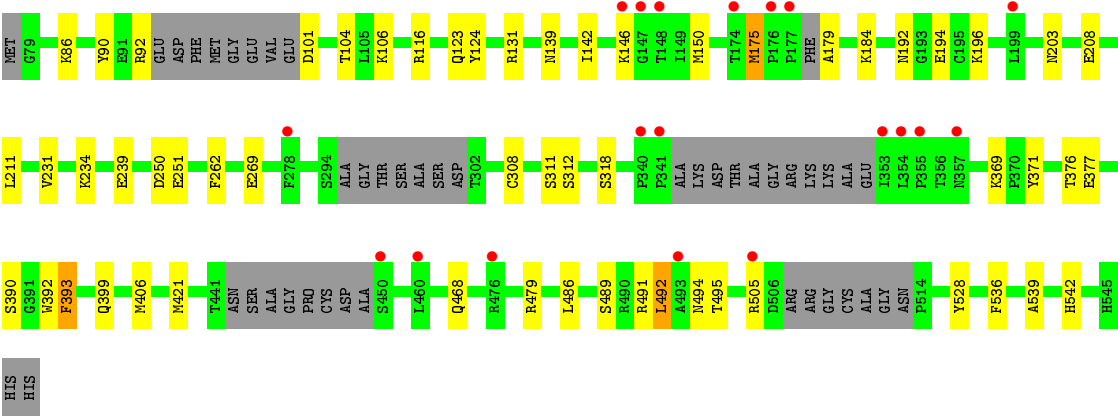
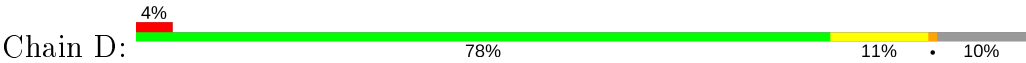


• Molecule 1: Tyrosyl-DNA phosphodiesterase 1





● Molecule 1: Tyrosyl-DNA phosphodiesterase 1



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	64.35Å 82.19Å 99.29Å 86.79° 85.57° 66.32°	Depositor
Resolution (Å)	50.00 – 2.00 36.24 – 1.67	Depositor EDS
% Data completeness (in resolution range)	97.2 (50.00-2.00) 82.9 (36.24-1.67)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.56 (at 1.67Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.174 , 0.213 0.175 , 0.212	Depositor DCC
R_{free} test set	8984 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å ²)	25.0	Xtriage
Anisotropy	0.143	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 41.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.007 for -h,-h+k,-l	Xtriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	14719	wwPDB-VP
Average B, all atoms (Å ²)	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.26% 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 [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: SO4

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	1.14	7/3598 (0.2%)	0.90	6/4871 (0.1%)
1	B	1.07	4/3566 (0.1%)	0.88	5/4826 (0.1%)
1	C	1.08	5/3602 (0.1%)	0.90	6/4875 (0.1%)
1	D	1.06	6/3564 (0.2%)	0.90	7/4822 (0.1%)
All	All	1.09	22/14330 (0.2%)	0.90	24/19394 (0.1%)

The worst 5 of 22 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	308	CYS	CB-SG	-15.38	1.56	1.82
1	A	312	SER	CB-OG	-8.72	1.30	1.42
1	C	308	CYS	CB-SG	-8.44	1.68	1.82
1	D	308	CYS	CB-SG	-7.84	1.69	1.82
1	B	308	CYS	CB-SG	-7.44	1.69	1.82

The worst 5 of 24 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	491	ARG	NE-CZ-NH2	-7.17	116.71	120.30
1	D	175	MET	CG-SD-CE	7.08	111.52	100.20
1	B	175	MET	CG-SD-CE	6.90	111.23	100.20
1	D	491	ARG	NE-CZ-NH2	-6.72	116.94	120.30
1	C	312	SER	CB-CA-C	-6.70	97.37	110.10

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	3502	0	3485	36	0
1	B	3469	0	3448	50	0
1	C	3503	0	3482	42	1
1	D	3466	0	3457	37	0
2	A	20	0	0	2	0
2	B	5	0	0	0	0
2	D	10	0	0	0	1
3	A	237	0	0	5	0
3	B	151	0	0	13	0
3	C	199	0	0	5	0
3	D	157	0	0	11	0
All	All	14719	0	13872	152	1

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 152 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:D:406[A]:MET:HE1	3:D:580:HOH:O	1.37	1.25
1:B:421[A]:MET:CE	3:B:571:HOH:O	1.85	1.21
1:C:302:THR:HG22	1:C:442:ASN:HB2	1.07	1.06
1:B:421[A]:MET:HE3	3:B:571:HOH:O	1.47	1.05
1:C:494:ASN:ND2	1:C:496:ARG:HE	1.57	1.01

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:495:THR:OG1	2:D:3:SO4:O2[1_554]	1.88	0.32

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	424/470 (90%)	414 (98%)	9 (2%)	1 (0%)	47	44
1	B	415/470 (88%)	404 (97%)	10 (2%)	1 (0%)	47	44
1	C	421/470 (90%)	410 (97%)	10 (2%)	1 (0%)	47	44
1	D	414/470 (88%)	405 (98%)	8 (2%)	1 (0%)	47	44
All	All	1674/1880 (89%)	1633 (98%)	37 (2%)	4 (0%)	47	44

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	211	LEU
1	B	211	LEU
1	D	211	LEU
1	C	211	LEU

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	394/423 (93%)	383 (97%)	11 (3%)	43	44
1	B	391/423 (92%)	382 (98%)	9 (2%)	50	53
1	C	394/423 (93%)	383 (97%)	11 (3%)	43	44
1	D	391/423 (92%)	382 (98%)	9 (2%)	50	53
All	All	1570/1692 (93%)	1530 (98%)	40 (2%)	47	49

5 of 40 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	468	GLN
1	C	150	MET
1	D	371	TYR
1	C	101	ASP
1	C	301	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 32 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	399	GLN
1	C	139	ASN
1	D	396	GLN
1	B	494	ASN
1	C	242	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

5.6 Ligand geometry ⓘ

7 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	$\# Z > 2$
2	SO4	B	5	-	4,4,4	0.40	0	6,6,6	0.19	0
2	SO4	D	1	-	4,4,4	0.20	0	6,6,6	0.24	0
2	SO4	A	2	-	4,4,4	0.20	0	6,6,6	0.19	0
2	SO4	A	6	-	4,4,4	0.23	0	6,6,6	0.38	0
2	SO4	D	3	-	4,4,4	0.33	0	6,6,6	0.66	0
2	SO4	A	7	-	4,4,4	0.24	0	6,6,6	0.20	0
2	SO4	A	4	-	4,4,4	0.66	0	6,6,6	0.68	0

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.

3 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	D	3	SO4	0	1
2	A	7	SO4	1	0
2	A	4	SO4	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	432/470 (91%)	-0.25	12 (2%) 53 51	21, 38, 77, 107	0
1	B	426/470 (90%)	-0.14	22 (5%) 27 26	22, 40, 77, 109	1 (0%)
1	C	430/470 (91%)	-0.25	13 (3%) 50 49	22, 40, 80, 113	1 (0%)
1	D	425/470 (90%)	-0.11	19 (4%) 33 32	23, 41, 79, 117	0
All	All	1713/1880 (91%)	-0.19	66 (3%) 39 38	21, 40, 79, 117	2 (0%)

The worst 5 of 66 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	341	PRO	5.7
1	D	278	PHE	5.4
1	A	341	PRO	5.3
1	D	354	LEU	5.3
1	D	353	ILE	5.3

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 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
2	SO4	B	5	5/5	0.69	0.15	142,142,144,145	5
2	SO4	A	2	5/5	0.75	0.34	160,160,162,162	0
2	SO4	D	1	5/5	0.76	0.24	131,132,133,133	5
2	SO4	A	7	5/5	0.85	0.31	141,142,144,144	0
2	SO4	A	6	5/5	0.90	0.21	99,100,100,101	5
2	SO4	A	4	5/5	0.92	0.11	77,79,82,82	0
2	SO4	D	3	5/5	0.94	0.25	131,138,138,139	0

6.5 Other polymers [i](#)

There are no such residues in this entry.