



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 9, 2017 – 09:35 PM EDT

PDB ID : 2O1X  
Title : 1-deoxy-D-xylulose 5-phosphate synthase (DXS) from *Deinococcus radiodurans*  
Authors : Xiang, S.; Usunow, G.; Lange, G.; Busch, M.; Tong, L.  
Deposited on : unknown  
Resolution : 2.90 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](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/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 : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20030345  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20030345

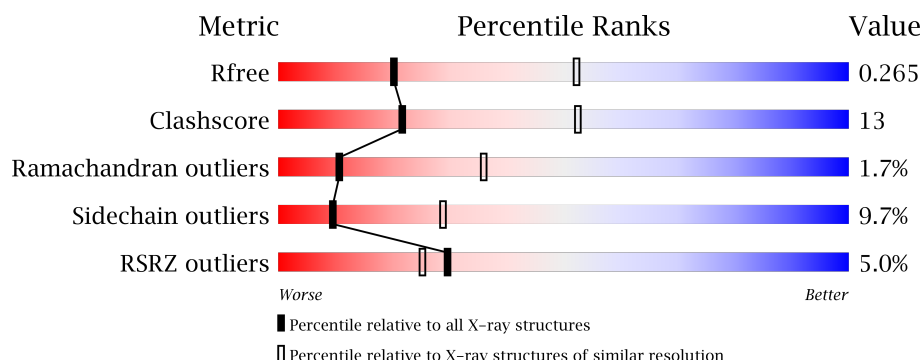
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.90 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))
$R_{free}$	100719	1586 (2.90-2.90)
Clashscore	112137	1807 (2.90-2.90)
Ramachandran outliers	110173	1768 (2.90-2.90)
Sidechain outliers	110143	1770 (2.90-2.90)
RSRZ outliers	101464	1596 (2.90-2.90)

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  $\geq 3$ , 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	A	629	<div> <div>8%</div> <div>65%</div> <div>23%</div> <div>8%</div> </div>
1	B	629	<div> <div>8%</div> <div>64%</div> <div>22%</div> <div>5%</div> <div>8%</div> </div>
1	C	629	<div> <div>8%</div> <div>58%</div> <div>23%</div> <div>14%</div> </div>
1	D	629	<div> <div>8%</div> <div>58%</div> <div>24%</div> <div>16%</div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	MG	A	2001	-	-	-	X
2	MG	B	2002	-	-	-	X

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 16910 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 1-deoxy-D-xylulose-5-phosphate synthase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	578	Total	C	N	O	S	0	0	0
			4362	2753	771	821	17			
1	B	579	Total	C	N	O	S	0	0	0
			4376	2761	775	823	17			
1	C	538	Total	C	N	O	S	0	0	0
			4063	2558	724	766	15			
1	D	530	Total	C	N	O	S	0	0	0
			4001	2521	715	750	15			

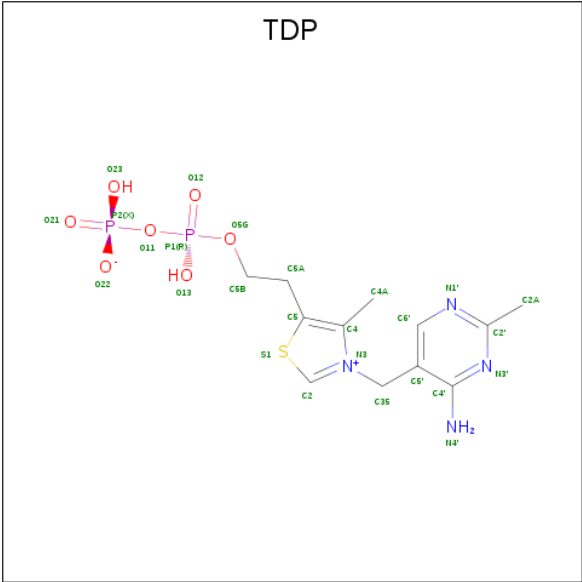
There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	130	THR	ALA	ENGINEERED	UNP Q9RUB5
B	130	THR	ALA	ENGINEERED	UNP Q9RUB5
C	130	THR	ALA	ENGINEERED	UNP Q9RUB5
D	130	THR	ALA	ENGINEERED	UNP Q9RUB5

- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	1	Total	Mg	0	0
			1	1		
2	A	1	Total	Mg	0	0
			1	1		
2	D	1	Total	Mg	0	0
			1	1		
2	C	1	Total	Mg	0	0
			1	1		

- Molecule 3 is THIAMIN DIPHOSPHATE (three-letter code: TDP) (formula: C<sub>12</sub>H<sub>18</sub>N<sub>4</sub>O<sub>7</sub>P<sub>2</sub>S).

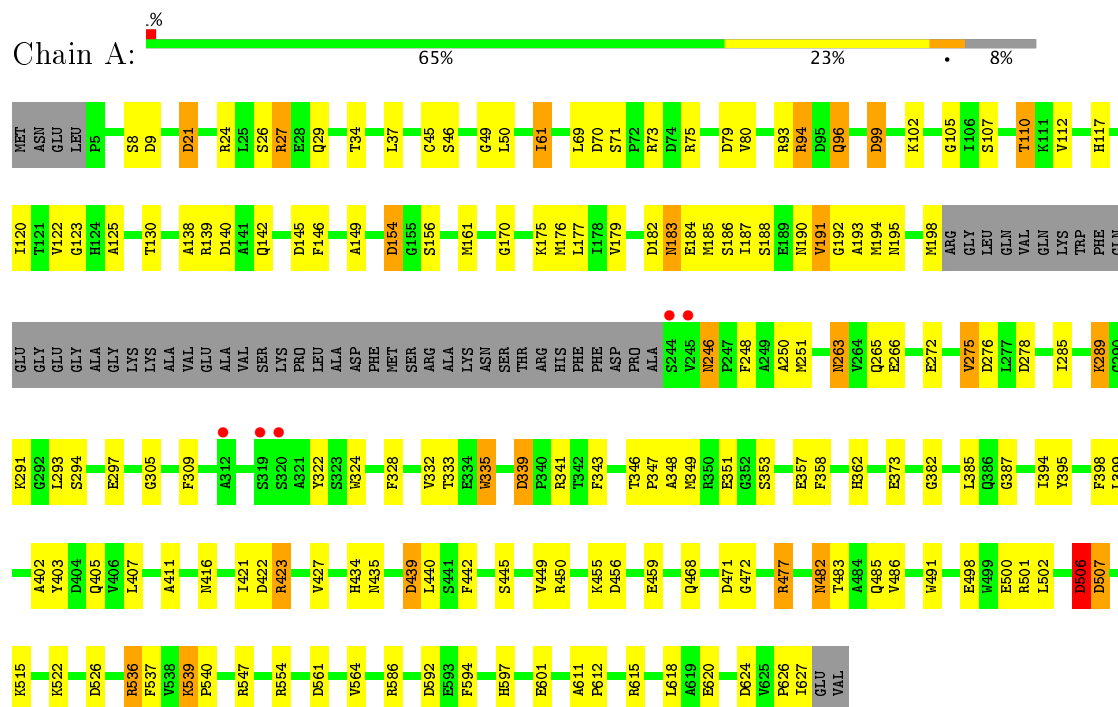


Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
3	A	1	Total	C	N	O	P	S	0	0
			26	12	4	7	2	1		
3	B	1	Total	C	N	O	P	S	0	0
			26	12	4	7	2	1		
3	C	1	Total	C	N	O	P	S	0	0
			26	12	4	7	2	1		
3	D	1	Total	C	N	O	P	S	0	0
			26	12	4	7	2	1		

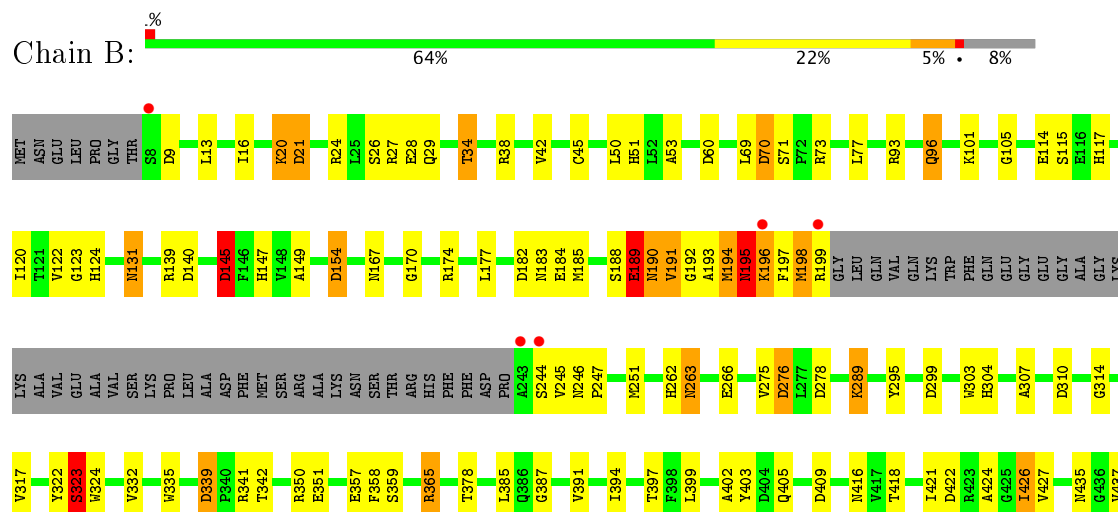
### 3 Residue-property plots

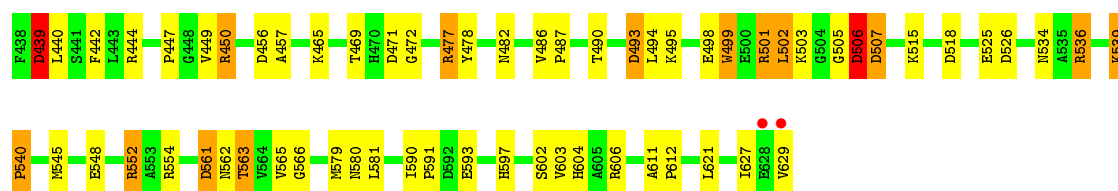
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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: 1-deoxy-D-xylulose-5-phosphate synthase

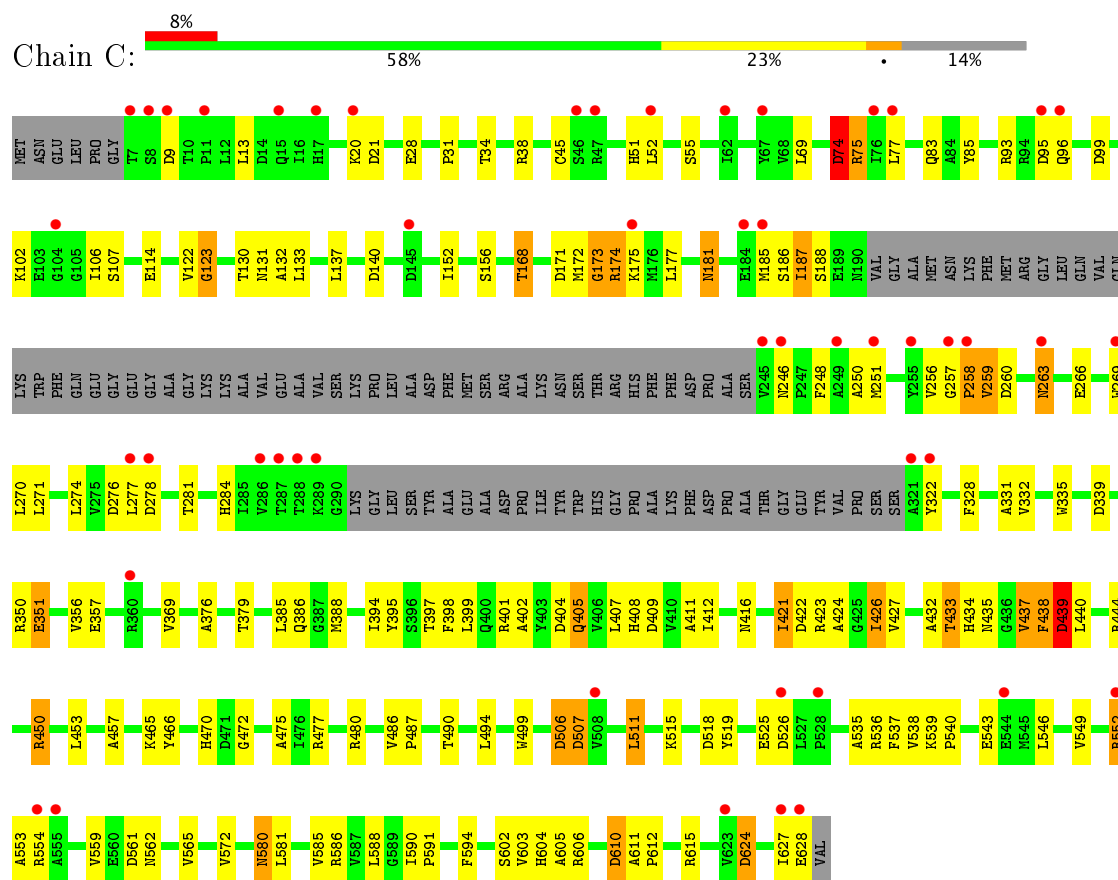


#### • Molecule 1: 1-deoxy-D-xylulose-5-phosphate synthase

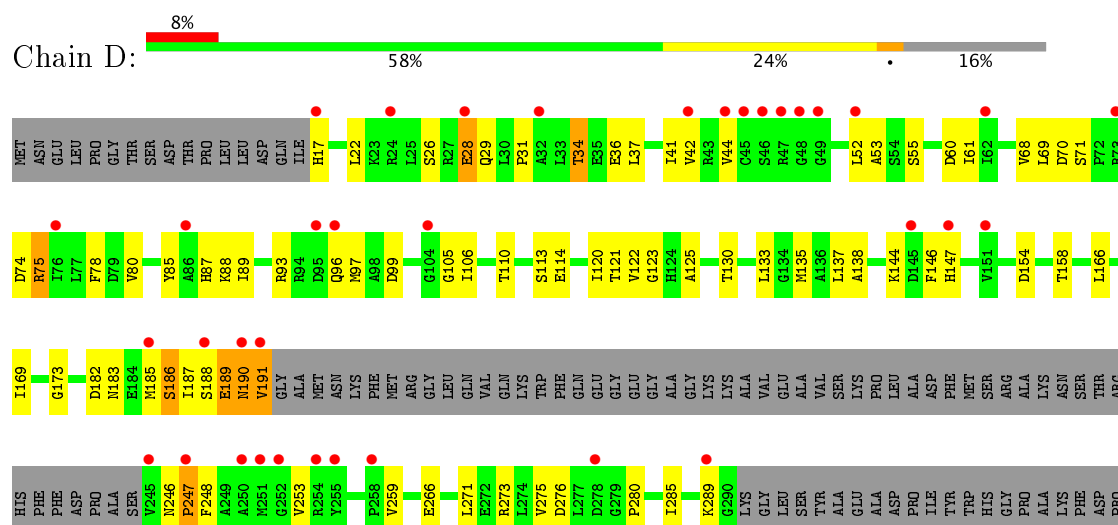


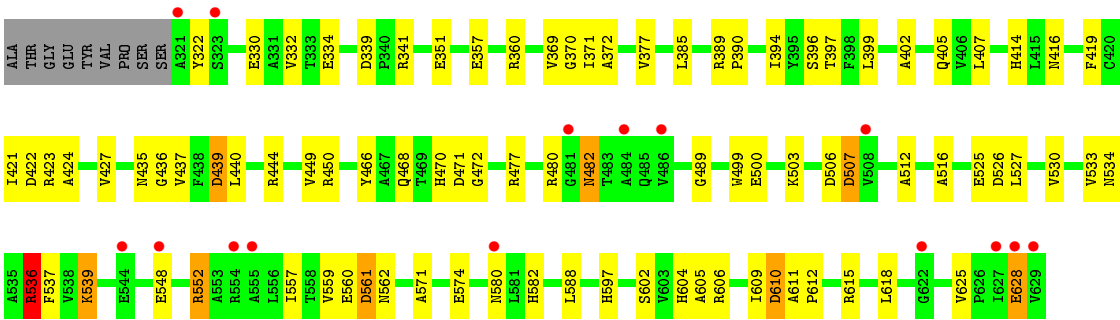


• Molecule 1: 1-deoxy-D-xylulose-5-phosphate synthase



• Molecule 1: 1-deoxy-D-xylulose-5-phosphate synthase







## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	78.26Å 154.06Å 124.88Å 90.00° 98.91° 90.00°	Depositor
Resolution (Å)	30.00 – 2.90 29.25 – 2.90	Depositor EDS
% Data completeness (in resolution range)	97.4 (30.00-2.90) 97.4 (29.25-2.90)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	0.11	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.44 (at 2.90Å)	Xtriage
Refinement program	REFMAC 5.2.0003	Depositor
R, $R_{free}$	0.209 , 0.272 0.205 , 0.265	Depositor DCC
$R_{free}$ test set	3204 reflections (5.09%)	DCC
Wilson B-factor (Å <sup>2</sup> )	40.7	Xtriage
Anisotropy	0.152	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 33.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	16910	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	39.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.83% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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: MG, TDP

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	0.53	0/4452	0.85	19/6049 (0.3%)
1	B	0.53	0/4465	0.87	16/6066 (0.3%)
1	C	0.43	0/4139	0.80	18/5621 (0.3%)
1	D	0.83	6/4076 (0.1%)	0.80	15/5534 (0.3%)
All	All	0.60	6/17132 (0.0%)	0.83	68/23270 (0.3%)

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
1	A	0	1
1	B	0	1
1	C	0	1
All	All	0	3

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	36	GLU	CD-OE2	28.08	1.56	1.25
1	D	628	GLU	CD-OE1	23.04	1.50	1.25
1	D	628	GLU	CD-OE2	18.05	1.45	1.25
1	D	36	GLU	CD-OE1	16.34	1.43	1.25
1	D	28	GLU	CD-OE2	9.40	1.35	1.25

The worst 5 of 68 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	182	ASP	CB-CG-OD2	7.46	125.02	118.30

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	561	ASP	CB-CG-OD2	7.41	124.97	118.30
1	B	310	ASP	CB-CG-OD2	7.27	124.84	118.30
1	A	154	ASP	CB-CG-OD2	7.21	124.79	118.30
1	C	422	ASP	CB-CG-OD2	7.20	124.78	118.30

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	191	VAL	Peptide
1	B	539	LYS	Peptide
1	C	438	PHE	Peptide

## 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	4362	0	4337	101	0
1	B	4376	0	4352	140	0
1	C	4063	0	4054	115	0
1	D	4001	0	3997	106	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
3	A	26	0	16	2	0
3	B	26	0	16	4	0
3	C	26	0	16	1	0
3	D	26	0	16	2	0
All	All	16910	0	16804	451	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 13.

The worst 5 of 451 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:C:1003:TDP:H2	3:C:1003:TDP:C2	0.97	1.49
3:B:1002:TDP:H2	3:B:1002:TDP:C2	0.97	1.47
3:D:1004:TDP:H2	3:D:1004:TDP:C2	0.97	1.47
3:A:1001:TDP:C2	3:A:1001:TDP:H2	0.97	1.46
1:D:439:ASP:OD2	1:D:477:ARG:HD2	1.47	1.11

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	A	574/629 (91%)	536 (93%)	34 (6%)	4 (1%)	25	60
1	B	575/629 (91%)	531 (92%)	32 (6%)	12 (2%)	8	30
1	C	532/629 (85%)	473 (89%)	48 (9%)	11 (2%)	8	30
1	D	524/629 (83%)	470 (90%)	44 (8%)	10 (2%)	9	33
All	All	2205/2516 (88%)	2010 (91%)	158 (7%)	37 (2%)	11	36

5 of 37 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	507	ASP
1	B	194	MET
1	B	195	ASN
1	B	244	SER
1	B	439	ASP

### 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	A	452/493 (92%)	404 (89%)	48 (11%)	8	24
1	B	453/493 (92%)	404 (89%)	49 (11%)	7	23
1	C	422/493 (86%)	380 (90%)	42 (10%)	9	27
1	D	414/493 (84%)	384 (93%)	30 (7%)	17	43
All	All	1741/1972 (88%)	1572 (90%)	169 (10%)	9	29

5 of 169 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	385	LEU
1	B	593	GLU
1	D	273	ARG
1	B	439	ASP
1	B	503	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 58 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	416	ASN
1	C	29	GLN
1	D	577	ASN
1	B	435	ASN
1	B	595	GLN

### 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

Of 8 ligands modelled in this entry, 4 are monoatomic - leaving 4 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
3	TDP	A	1001	2	21,27,27	1.76	4 (19%)	25,40,40	1.94	9 (36%)
3	TDP	B	1002	2	21,27,27	1.67	4 (19%)	25,40,40	1.85	7 (28%)
3	TDP	C	1003	2	21,27,27	1.84	7 (33%)	25,40,40	1.95	6 (24%)
3	TDP	D	1004	2	21,27,27	1.69	6 (28%)	25,40,40	1.84	8 (32%)

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	TDP	A	1001	2	-	0/16/17/17	0/2/2/2
3	TDP	B	1002	2	-	0/16/17/17	0/2/2/2
3	TDP	C	1003	2	-	0/16/17/17	0/2/2/2
3	TDP	D	1004	2	-	0/16/17/17	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	B	1002	TDP	C4-N3	-4.18	1.36	1.39
3	A	1001	TDP	C4-N3	-3.11	1.37	1.39
3	D	1004	TDP	C4-N3	-3.06	1.37	1.39
3	C	1003	TDP	C4-N3	-2.30	1.37	1.39
3	C	1003	TDP	P2-O22	2.04	1.63	1.54

The worst 5 of 30 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	1001	TDP	C4A-C4-C5	-3.85	119.83	127.29
3	C	1003	TDP	C4A-C4-C5	-3.36	120.78	127.29

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	1002	TDP	N1'-C2'-N3'	-3.30	119.88	125.59
3	A	1001	TDP	N1'-C2'-N3'	-3.20	120.05	125.59
3	D	1004	TDP	N1'-C2'-N3'	-3.09	120.24	125.59

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

4 monomers are involved in 9 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	1001	TDP	2	0
3	B	1002	TDP	4	0
3	C	1003	TDP	1	0
3	D	1004	TDP	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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	578/629 (91%)	-0.48	5 (0%) 84 83	-5, 10, 34, 53	0
1	B	579/629 (92%)	-0.48	7 (1%) 79 77	-3, 11, 39, 75	0
1	C	538/629 (85%)	0.37	49 (9%) 10 7	6, 65, 139, 191	0
1	D	530/629 (84%)	0.34	51 (9%) 9 6	7, 60, 133, 175	0
All	All	2225/2516 (88%)	-0.08	112 (5%) 30 25	-5, 24, 121, 191	0

The worst 5 of 112 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	15	GLN	6.1
1	D	46	SER	6.0
1	D	49	GLY	5.6
1	C	321	ALA	5.5
1	C	628	GLU	5.5

### 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	MG	B	2002	1/1	0.93	0.21	3.02	8,8,8,8	0
2	MG	A	2001	1/1	0.97	0.21	2.41	9,9,9,9	0
3	TDP	B	1002	26/26	0.98	0.13	-0.00	2,3,4,6	0
3	TDP	D	1004	26/26	0.92	0.16	-0.48	40,44,48,48	0
3	TDP	C	1003	26/26	0.93	0.16	-0.66	50,50,52,53	0
3	TDP	A	1001	26/26	0.98	0.11	-0.83	2,2,4,5	0
2	MG	D	2004	1/1	0.93	0.12	-0.96	35,35,35,35	0
2	MG	C	2003	1/1	0.93	0.12	-1.67	33,33,33,33	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.