



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 05:18 AM GMT

PDB ID : 2Q5J  
Title : X-ray structure of phenylpyruvate decarboxylase in complex with 3-deaza-ThDP  
Authors : Versees, W.; Spaepen, S.; Wood, M.D.; Leeper, F.J.; Vanderleyden, J.; Steyaert, J.  
Deposited on : 2007-06-01  
Resolution : 3.20 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](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 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
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) : trunk26865

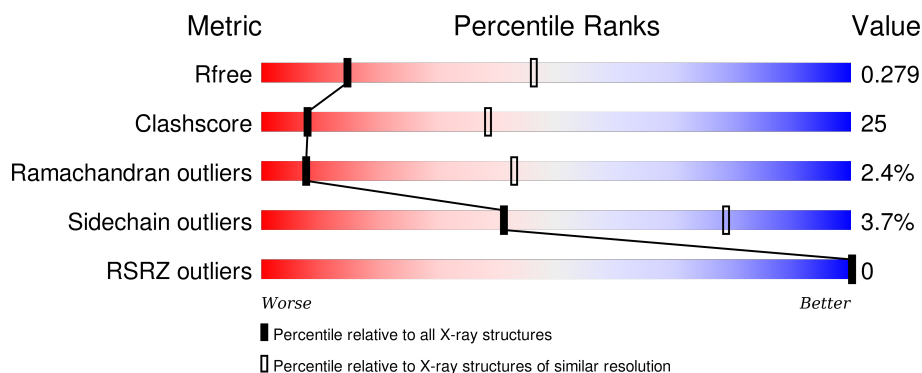
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.20 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))
$R_{free}$	91344	1124 (3.24-3.16)
Clashscore	102246	1024 (3.22-3.18)
Ramachandran outliers	100387	1004 (3.22-3.18)
Sidechain outliers	100360	1003 (3.22-3.18)
RSRZ outliers	91569	1129 (3.24-3.16)

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	565	 52% 38% • 7%
1	B	565	 55% 36% •• 6%

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 8008 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 Phenylpyruvate decarboxylase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	523	Total	C	N	O	S	0	0	0
			3884	2459	678	724	23			
1	B	530	Total	C	N	O	S	0	0	0
			3935	2488	687	737	23			

There are 44 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-19	MET	-	EXPRESSION TAG	UNP P51852
A	-18	GLY	-	EXPRESSION TAG	UNP P51852
A	-17	SER	-	EXPRESSION TAG	UNP P51852
A	-16	SER	-	EXPRESSION TAG	UNP P51852
A	-15	HIS	-	EXPRESSION TAG	UNP P51852
A	-14	HIS	-	EXPRESSION TAG	UNP P51852
A	-13	HIS	-	EXPRESSION TAG	UNP P51852
A	-12	HIS	-	EXPRESSION TAG	UNP P51852
A	-11	HIS	-	EXPRESSION TAG	UNP P51852
A	-10	HIS	-	EXPRESSION TAG	UNP P51852
A	-9	SER	-	EXPRESSION TAG	UNP P51852
A	-8	SER	-	EXPRESSION TAG	UNP P51852
A	-7	GLY	-	EXPRESSION TAG	UNP P51852
A	-6	LEU	-	EXPRESSION TAG	UNP P51852
A	-5	VAL	-	EXPRESSION TAG	UNP P51852
A	-4	PRO	-	EXPRESSION TAG	UNP P51852
A	-3	ARG	-	EXPRESSION TAG	UNP P51852
A	-2	GLY	-	EXPRESSION TAG	UNP P51852
A	-1	SER	-	EXPRESSION TAG	UNP P51852
A	0	HIS	-	EXPRESSION TAG	UNP P51852
A	155	GLN	LEU	EXPRESSION TAG	UNP P51852
A	327	ARG	GLY	SEE REMARK 999	UNP P51852
B	-19	MET	-	EXPRESSION TAG	UNP P51852
B	-18	GLY	-	EXPRESSION TAG	UNP P51852
B	-17	SER	-	EXPRESSION TAG	UNP P51852

*Continued on next page...*

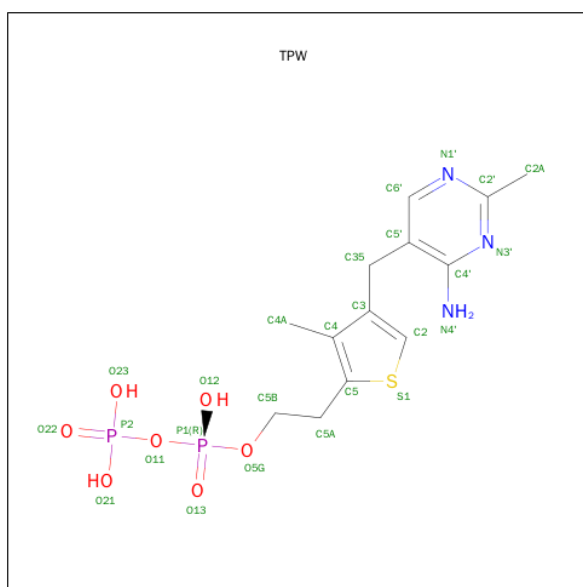
*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	-16	SER	-	EXPRESSION TAG	UNP P51852
B	-15	HIS	-	EXPRESSION TAG	UNP P51852
B	-14	HIS	-	EXPRESSION TAG	UNP P51852
B	-13	HIS	-	EXPRESSION TAG	UNP P51852
B	-12	HIS	-	EXPRESSION TAG	UNP P51852
B	-11	HIS	-	EXPRESSION TAG	UNP P51852
B	-10	HIS	-	EXPRESSION TAG	UNP P51852
B	-9	SER	-	EXPRESSION TAG	UNP P51852
B	-8	SER	-	EXPRESSION TAG	UNP P51852
B	-7	GLY	-	EXPRESSION TAG	UNP P51852
B	-6	LEU	-	EXPRESSION TAG	UNP P51852
B	-5	VAL	-	EXPRESSION TAG	UNP P51852
B	-4	PRO	-	EXPRESSION TAG	UNP P51852
B	-3	ARG	-	EXPRESSION TAG	UNP P51852
B	-2	GLY	-	EXPRESSION TAG	UNP P51852
B	-1	SER	-	EXPRESSION TAG	UNP P51852
B	0	HIS	-	EXPRESSION TAG	UNP P51852
B	155	GLN	LEU	EXPRESSION TAG	UNP P51852
B	327	ARG	GLY	SEE REMARK 999	UNP P51852

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

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

- Molecule 3 is 2-{4-[(4-AMINO-2-METHYLPYRIMIDIN-5-YL)METHYL]-3-METHYLTHIOPHEN-2-YL}ETHYL TRIHYDROGEN DIPHOSPHATE (three-letter code: TPW) (formula: C<sub>13</sub>H<sub>19</sub>N<sub>3</sub>O<sub>7</sub>P<sub>2</sub>S).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
3	B	1	Total 26	C 13	N 3	O 7	P 2	S 1	0	0
3	A	1	Total 26	C 13	N 3	O 7	P 2	S 1	0	0

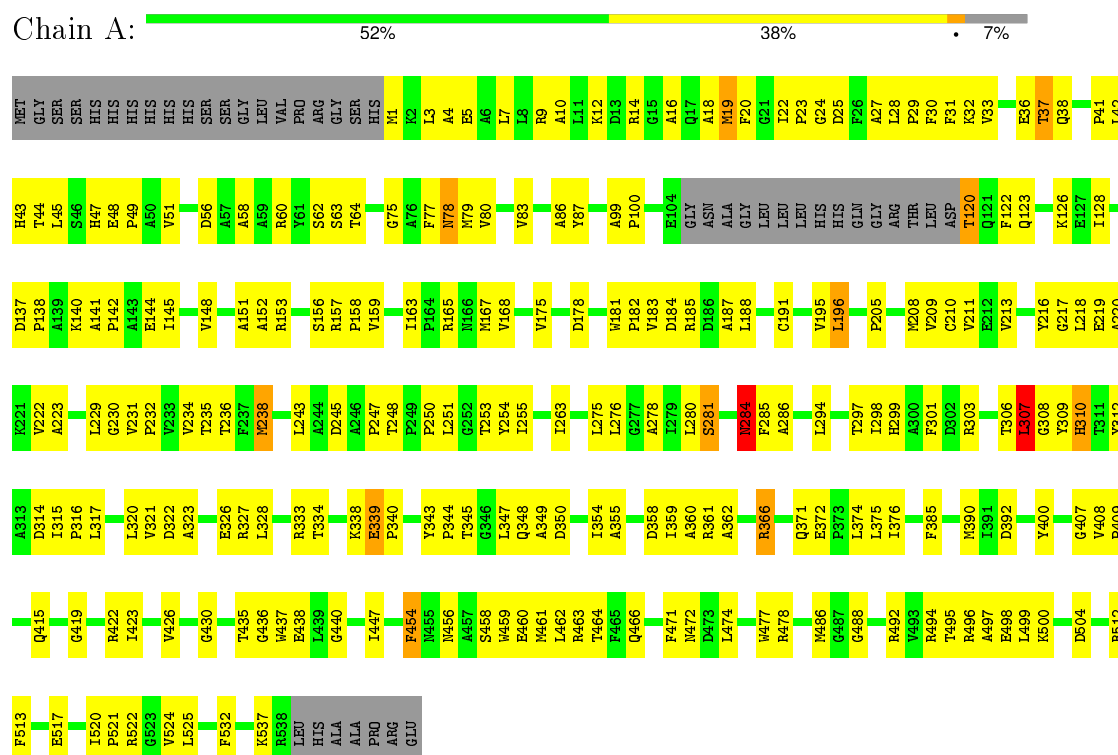
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	56	Total O 56 56	0	0
4	B	79	Total O 79 79	0	0

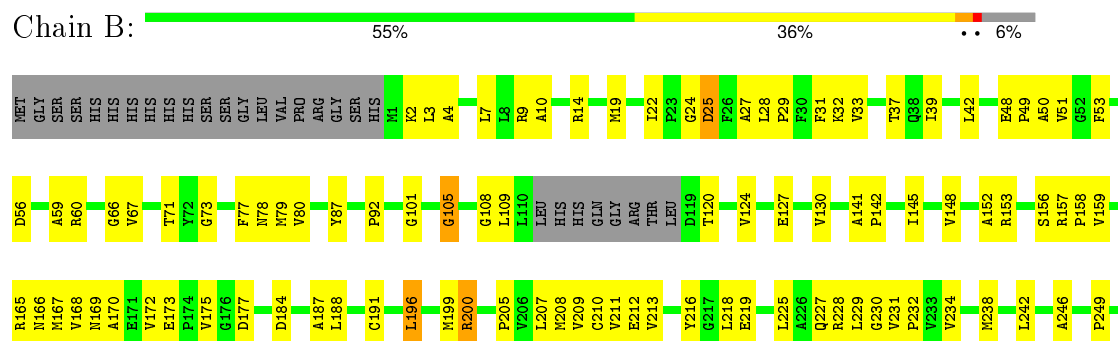
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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 ( $\text{RSRZ} > 2$ ). 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: Phenylpyruvate decarboxylase



#### • Molecule 1: Phenylpyruvate decarboxylase



V257	Q348	H437	P521
A258	A349	E438	LEU
G259	D350	L439	HIS
D260	G351	G440	ALA
E261	E352	R444	ALA
E262	P353	I447	PRO
I263	I354	D448	ARG
T264	I359	P449	GLU
D271	R360	I450	
L275	A361	V451	
L276	A362	I452	
G277	V363	L453	
A278	R366	F454	
I279	V367	A457	
L280	R368	S458	
S281	Q371	V459	
D282	E372	E460	
T283	P373	M461	
I284	L374	L462	
Q289	L375	R463	
	I376	T464	
K296	M380	F465	
T297	M388	Q466	
I298		S469	
F301	A393	A470	
T306	G394	F471	
I307	L395	N472	
G308	Y400	D473	
Y309	Y401	L474	
H310	A402	N477	
T311	G403	R478	
Y312	M404	F479	
A313	G405	M482	
D314	F406		
I315	G407	M486	
P316	V408	Q487	
L317	P409	V491	
		R492	
V321	A414	T495	
D322	Q415	R496	
A323	S418	A497	
R327		E498	
	K421	L499	
P330	M422	K500	
R333	I423		
	L424	D504	
G337	T425		
K338		R510	
P339	F432	G511	
P340	Q433	R512	
H341	M434		
	T435	E517	
	G436		

## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	100.54Å 179.81Å 121.03Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 3.20 36.81 – 3.20	Depositor EDS
% Data completeness (in resolution range)	99.4 (50.00-3.20) 96.7 (36.81-3.20)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.16	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.55 (at 3.18Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.176 , 0.289 0.171 , 0.279	Depositor DCC
$R_{free}$ test set	911 reflections (5.10%)	DCC
Wilson B-factor (Å <sup>2</sup> )	62.0	Xtriage
Anisotropy	0.146	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 46.6	EDS
Estimated twinning fraction	0.013 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.030 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 17865 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	8008	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	41.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.53% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: TPW, MG

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	0.35	0/3959	0.62	0/5383
1	B	0.36	0/4010	0.63	1/5452 (0.0%)
All	All	0.35	0/7969	0.62	1/10835 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	105	GLY	N-CA-C	7.99	133.06	113.10

There are no chirality outliers.

There are no planarity outliers.

### 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	A	3884	0	3876	202	0
1	B	3935	0	3921	203	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	26	0	16	2	0
3	B	26	0	16	3	0
4	A	56	0	0	6	0
4	B	79	0	0	2	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	8008	0	7829	392	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 25.

The worst 5 of 392 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:18:ALA:HA	1:A:41:PRO:HD2	1.41	1.00
1:A:120:THR:HG22	1:A:123:GLN:H	1.35	0.91
1:A:14:ARG:HD3	1:A:153:ARG:HD2	1.54	0.89
1:A:366:ARG:HB3	1:A:371:GLN:HE21	1.36	0.89
1:A:458:SER:HA	1:A:474:LEU:HG	1.55	0.88

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	519/565 (92%)	443 (85%)	64 (12%)	12 (2%)	8	44
1	B	526/565 (93%)	439 (84%)	74 (14%)	13 (2%)	7	41
All	All	1045/1130 (92%)	882 (84%)	138 (13%)	25 (2%)	7	43

5 of 25 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	105	GLY
1	B	340	PRO
1	B	349	ALA
1	B	407	GLY

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	38	GLN

### 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	387/426 (91%)	371 (96%)	16 (4%)	37	76
1	B	392/426 (92%)	379 (97%)	13 (3%)	45	81
All	All	779/852 (91%)	750 (96%)	29 (4%)	41	79

5 of 29 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	461	MET
1	B	196	LEU
1	B	471	PHE
1	A	478	ARG
1	B	200	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	371	GLN
1	A	472	ASN
1	B	371	GLN
1	A	299	HIS
1	B	289	GLN

### 5.3.3 RNA ⓘ

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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 4 ligands modelled in this entry, 2 are monoatomic - leaving 2 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	TPW	A	1002	2	22,27,27	1.80	5 (22%)	31,40,40	2.65	8 (25%)
3	TPW	B	1001	2	22,27,27	1.54	4 (18%)	31,40,40	2.73	9 (29%)

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	TPW	A	1002	2	-	0/15/17/17	0/2/2/2
3	TPW	B	1001	2	-	0/15/17/17	0/2/2/2

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	1002	TPW	C2-C3	-5.24	1.33	1.37
3	B	1001	TPW	C2-C3	-3.62	1.34	1.37
3	B	1001	TPW	C2-S1	2.03	1.73	1.70
3	A	1002	TPW	C4'-N3'	2.05	1.38	1.35
3	A	1002	TPW	C6'-C5'	2.22	1.42	1.37

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	1001	TPW	C3-C2-S1	-9.26	102.39	112.26
3	A	1002	TPW	C3-C2-S1	-8.72	102.97	112.26
3	B	1001	TPW	P1-O11-P2	-6.22	111.81	132.67
3	A	1002	TPW	P1-O11-P2	-6.00	112.54	132.67
3	B	1001	TPW	N1'-C2'-N3'	-2.43	121.10	125.60

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	1002	TPW	2	0
3	B	1001	TPW	3	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	523/565 (92%)	-0.60	0 100 100	10, 38, 66, 94	0
1	B	530/565 (93%)	-0.55	0 100 100	9, 39, 77, 99	0
All	All	1053/1130 (93%)	-0.58	0 100 100	9, 38, 72, 99	0

There are no RSRZ outliers to report.

### 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(Å <sup>2</sup> )	Q<0.9
2	MG	A	2002	1/1	0.98	0.20	0.98	50,50,50,50	0
3	TPW	A	1002	26/26	0.98	0.16	0.06	10,23,36,41	0
3	TPW	B	1001	26/26	0.98	0.16	-0.08	9,32,47,50	0
2	MG	B	2001	1/1	0.98	0.14	-0.17	22,22,22,22	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.