



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 14, 2017 – 08:44 AM EDT

PDB ID : 5TTH  
Title : Heterodimeric SpyCatcher/SpyTag-fused zebrafish TRAP1 in ATP/ADP-hybrid state  
Authors : Elnatan, D.; Betegon, M.; Liu, Y.; Agard, D.A.; Northeast Structural Genomics Consortium (NESG)  
Deposited on : unknown  
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	:	1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20029824
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-20029824

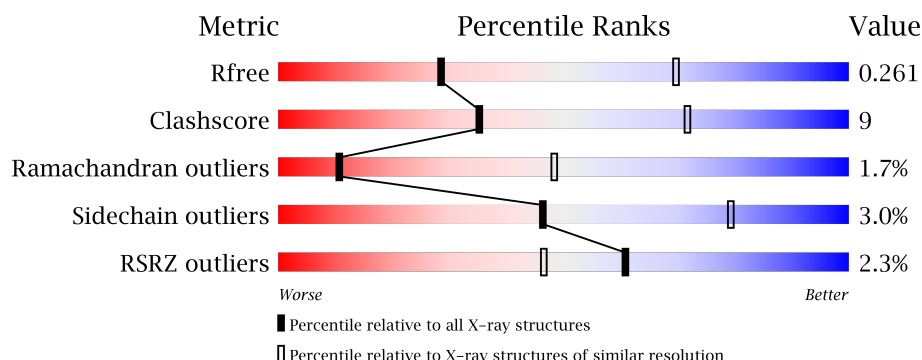
# 1 Overall quality at a glance ⓘ

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))
$R_{free}$	100719	1015 (3.22-3.18)
Clashscore	112137	1009 (3.20-3.20)
Ramachandran outliers	110173	1118 (3.22-3.18)
Sidechain outliers	110143	1117 (3.22-3.18)
RSRZ outliers	101464	1020 (3.22-3.18)

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	738	<div> <div>4%</div> <div> <div></div> <div>76%</div> <div>14%</div> <div>• 9%</div> </div> </div>
2	B	672	<div> <div>%</div> <div> <div></div> <div>75%</div> <div>15%</div> <div>•• 8%</div> </div> </div>

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 10447 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 C-terminal SpyCatcher fusion of wildtype zebrafish TNF receptor-associated protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	671	Total	C	N	O	S	0	2	0
			5371	3399	923	1026	23			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	67	GLY	-	expression tag	UNP A8WFFV1
A	68	ILE	-	expression tag	UNP A8WFFV1
A	69	ASP	-	expression tag	UNP A8WFFV1
A	70	PRO	-	expression tag	UNP A8WFFV1
A	71	PHE	-	expression tag	UNP A8WFFV1
A	72	THR	-	expression tag	UNP A8WFFV1
A	720	GLY	-	linker	UNP A8WFFV1
A	721	SER	-	linker	UNP A8WFFV1
A	734	GLU	ILE	engineered mutation	UNP Q8G9G1
A	769	TYR	MET	engineered mutation	UNP Q8G9G1

- Molecule 2 is a protein called C-terminal Spytag fusion of R417A zebrafish TNF receptor-associated protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	621	Total	C	N	O	S	0	4	0
			5012	3182	862	945	23			

There are 13 discrepancies between the modelled and reference sequences:

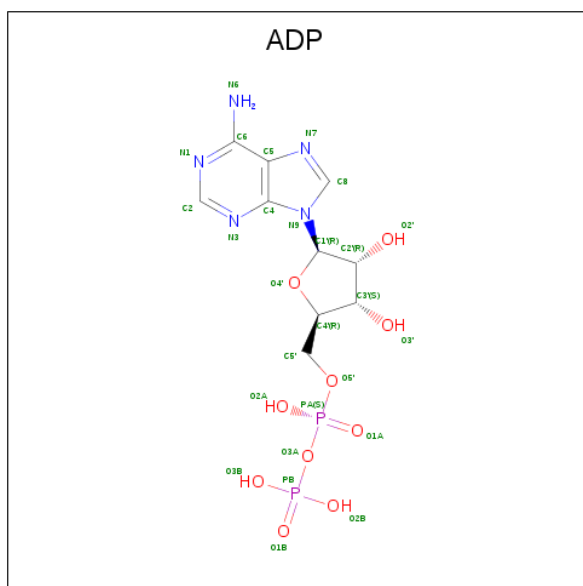
Chain	Residue	Modelled	Actual	Comment	Reference
B	67	GLY	-	expression tag	UNP A8WFFV1
B	68	ILE	-	expression tag	UNP A8WFFV1
B	69	ASP	-	expression tag	UNP A8WFFV1
B	70	PRO	-	expression tag	UNP A8WFFV1

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	71	PHE	-	expression tag	UNP A8WV1
B	72	THR	-	expression tag	UNP A8WV1
B	417	ALA	ARG	engineered mutation	UNP A8WV1
B	720	GLY	-	linker	UNP A8WV1
B	721	GLY	-	linker	UNP A8WV1
B	722	SER	-	linker	UNP A8WV1
B	723	GLY	-	linker	UNP A8WV1
B	724	SER	-	linker	UNP A8WV1
B	725	SER	-	linker	UNP A8WV1

- Molecule 3 is ADENOSINE-5'-DIPHOSPHATE (three-letter code: ADP) (formula:  $C_{10}H_{15}N_5O_{10}P_2$ ).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	A	1	Total	C	N	O	P	0	0
			27	10	5	10	2		
3	B	1	Total	C	N	O	P	0	0
			27	10	5	10	2		

- Molecule 4 is BERYLLIUM TRIFLUORIDE ION (three-letter code: BEF) (formula:  $BeF_3$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	Be	F	0	0
			4	1	3		
4	B	1	Total	Be	F	0	0
			4	1	3		

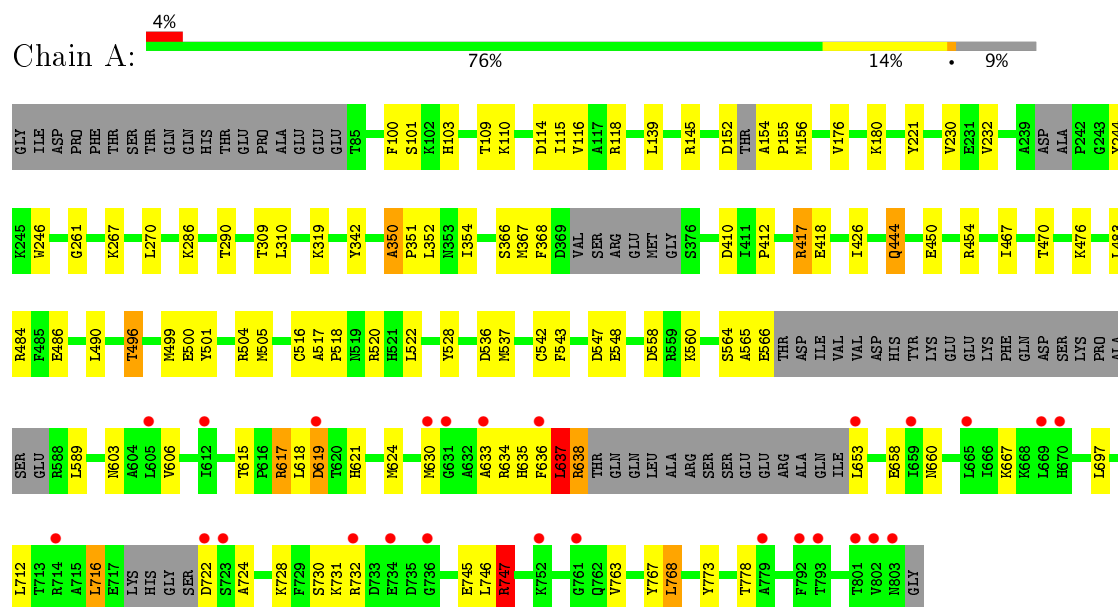
- Molecule 5 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

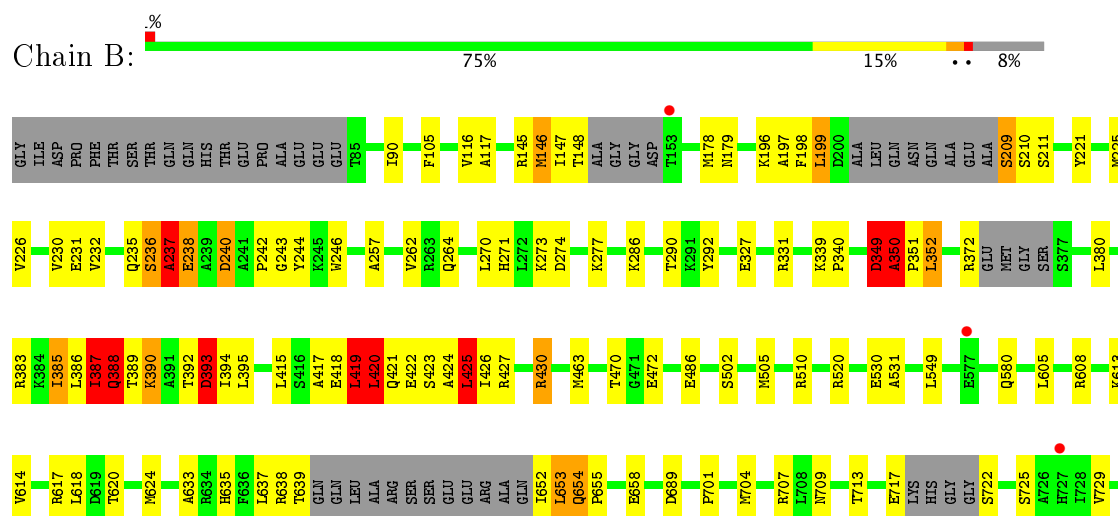
### 3 Residue-property plots

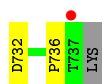
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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: C-terminal SpyCatcher fusion of wildtype zebrafish TNF receptor-associated protein 1



- Molecule 2: C-terminal Spytag fusion of R417A zebrafish TNF receptor-associated protein 1





## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	180.09 Å   95.71 Å   126.56 Å 90.00°   134.59°   90.00°	Depositor
Resolution (Å)	49.00 – 3.20 49.09 – 3.20	Depositor EDS
% Data completeness (in resolution range)	83.0 (49.00-3.20) 83.0 (49.09-3.20)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.15	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.49 (at 3.19 Å)	Xtriage
Refinement program	REFMAC 5.8.0155	Depositor
R, $R_{free}$	0.228 , 0.261 0.227 , 0.261	Depositor DCC
$R_{free}$ test set	1014 reflections (5.04%)	DCC
Wilson B-factor (Å <sup>2</sup> )	74.7	Xtriage
Anisotropy	0.145	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 79.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.45$ , $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	0.015 for h+2*l,k,-h-l 0.028 for h,-k,-h-l 0.039 for -h-2*l,-k,l	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	10447	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	126.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.21% 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, BEF, ADP

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	0.61	1/5473 (0.0%)	0.79	6/7377 (0.1%)
2	B	0.63	1/5116 (0.0%)	0.80	14/6895 (0.2%)
All	All	0.62	2/10589 (0.0%)	0.79	20/14272 (0.1%)

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	B	0	1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	393	ASP	CB-CG	-5.77	1.39	1.51
1	A	261	GLY	N-CA	-5.40	1.38	1.46

The worst 5 of 20 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	417	ARG	NE-CZ-NH2	15.85	128.22	120.30
1	A	417	ARG	NE-CZ-NH1	-10.06	115.27	120.30
2	B	349	ASP	CB-CG-OD1	-8.49	110.66	118.30
2	B	393	ASP	CB-CG-OD2	-7.42	111.62	118.30
2	B	352	LEU	CB-CG-CD1	-7.05	99.01	111.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	393	ASP	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	5371	0	5343	90	2
2	B	5012	0	5028	111	0
3	A	27	0	12	0	0
3	B	27	0	12	0	0
4	A	4	0	0	0	0
4	B	4	0	0	0	0
5	A	1	0	0	0	0
5	B	1	0	0	0	0
All	All	10447	0	10395	193	2

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 9.

The worst 5 of 193 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:745:GLU:OE1	1:A:747:ARG:NH2	1.61	1.31
2:B:419:LEU:HD21	2:B:422:GLU:OE1	1.29	1.29
2:B:419:LEU:CD2	2:B:420:LEU:HD22	1.64	1.27
1:A:712:LEU:O	1:A:716:LEU:HD12	1.28	1.27
2:B:385:ILE:O	2:B:387:ILE:CD1	1.83	1.25

All (2) 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:A:155:PRO:CD	1:A:778:THR:O[1_454]	1.76	0.44
1:A:155:PRO:CG	1:A:778:THR:O[1_454]	1.83	0.37

## 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	659/738 (89%)	631 (96%)	24 (4%)	4 (1%)	28	72
2	B	613/672 (91%)	568 (93%)	27 (4%)	18 (3%)	5	33
All	All	1272/1410 (90%)	1199 (94%)	51 (4%)	22 (2%)	11	48

5 of 22 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	350	ALA
2	B	198	PHE
2	B	210	SER
2	B	350	ALA
2	B	394	ILE

### 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	586/643 (91%)	574 (98%)	12 (2%)	60	86
2	B	553/587 (94%)	531 (96%)	22 (4%)	36	73
All	All	1139/1230 (93%)	1105 (97%)	34 (3%)	46	79

5 of 34 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	238	GLU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
2	B	383	ARG
2	B	614	VAL
2	B	349	ASP
1	A	637	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
2	B	97	GLN
2	B	235	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 6 ligands modelled in this entry, 2 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	ADP	A	901	5,4	25,29,29	1.02	1 (4%)	24,45,45	1.77	3 (12%)
4	BEF	A	902	3	0,3,3	0.00	-	0,3,3	0.00	-

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
3	ADP	B	801	5,4	25,29,29	1.01	2 (8%)	24,45,45	1.71	2 (8%)
4	BEF	B	802	3	0,3,3	0.00	-	0,3,3	0.00	-

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	ADP	A	901	5,4	-	0/12/32/32	0/3/3/3
4	BEF	A	902	3	-	0/0/0/0	0/0/0/0
3	ADP	B	801	5,4	-	0/12/32/32	0/3/3/3
4	BEF	B	802	3	-	0/0/0/0	0/0/0/0

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	801	ADP	C2-N3	2.29	1.36	1.32
3	A	901	ADP	C5-C4	2.50	1.46	1.40
3	B	801	ADP	C5-C4	2.74	1.46	1.40

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	901	ADP	N3-C2-N1	-6.79	122.94	128.86
3	B	801	ADP	N3-C2-N1	-6.09	123.55	128.86
3	B	801	ADP	C4-C5-N7	-2.96	106.55	109.41
3	A	901	ADP	C4-C5-N7	-2.72	106.78	109.41
3	A	901	ADP	O2A-PA-O1A	2.02	122.73	112.28

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues ⓘ

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	671/738 (90%)	0.20	26 (3%) 40 26	63, 134, 204, 250	0
2	B	621/672 (92%)	-0.13	4 (0%) 89 83	57, 114, 176, 281	0
All	All	1292/1410 (91%)	0.04	30 (2%) 61 46	57, 124, 196, 281	0

The worst 5 of 30 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	803	ASN	5.1
2	B	737	THR	5.0
1	A	723	SER	4.5
1	A	633	ALA	4.3
1	A	779	ALA	3.7

### 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
5	MG	A	903	1/1	0.97	0.20	0.45	48,48,48,48	0
4	BEF	B	802	4/4	0.95	0.14	-1.22	55,55,55,58	0
3	ADP	A	901	27/27	0.96	0.15	-1.38	50,57,66,70	0
5	MG	B	803	1/1	0.96	0.11	-1.42	38,38,38,38	0
4	BEF	A	902	4/4	0.95	0.15	-1.46	57,64,66,72	0
3	ADP	B	801	27/27	0.96	0.14	-1.61	47,54,74,79	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.