



# Full wwPDB X-ray Structure Validation Report ⓘ

Jan 23, 2017 – 09:18 PM EST

PDB ID : 5HSR  
Title : Fluorine substituted 5-methyl-6-(3',4'-difluoromethoxyphenylthio)thieno[2,3-d  
|pyrimidine-2,4-diamine  
Authors : Cody, V.  
Deposited on : 2016-01-26  
Resolution : 1.21 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 : 1.7.1 (RC1), CSD as537be (2016)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20028442  
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) : rb-20028442

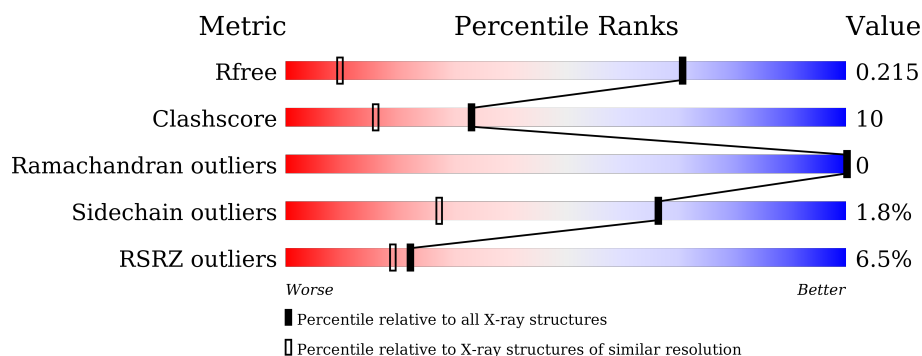
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.21 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	1197 (1.26-1.18)
Clashscore	102246	1295 (1.26-1.18)
Ramachandran outliers	100387	1239 (1.26-1.18)
Sidechain outliers	100360	1237 (1.26-1.18)
RSRZ outliers	91569	1201 (1.26-1.18)

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	186	

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 criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	63Y	A	202	-	-	X	-

## 2 Entry composition [i](#)

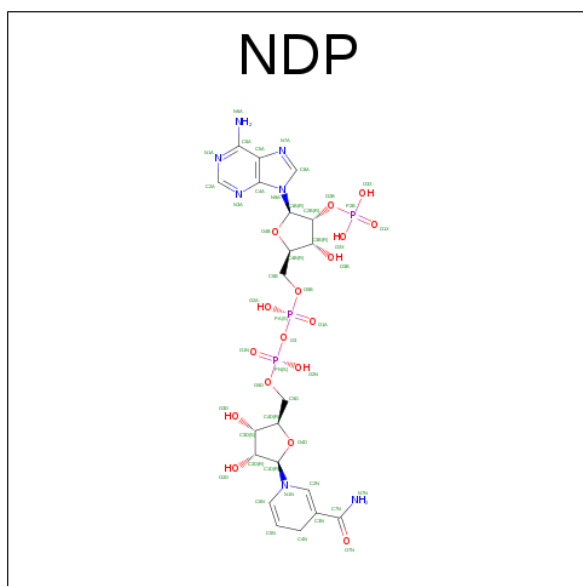
There are 4 unique types of molecules in this entry. The entry contains 1822 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Dihydrofolate reductase.

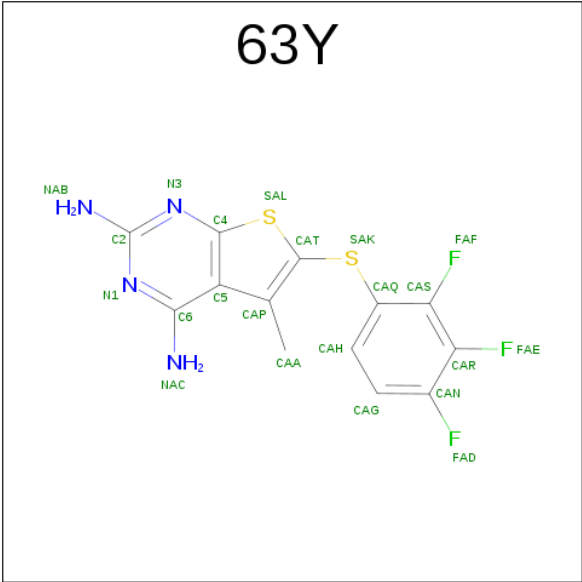
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	186	1544	990	258	287	9	0	6	0

- Molecule 2 is NADPH DIHYDRO-NICOTINAMIDE-ADENINE-DINUCLEOTIDE PHOSPHATE (three-letter code: NDP) (formula: C<sub>21</sub>H<sub>30</sub>N<sub>7</sub>O<sub>17</sub>P<sub>3</sub>).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
2	A	1	48	21	7	17	3	0	0

- Molecule 3 is 5-methyl-6-[(2,3,4-trifluorophenyl)sulfanyl]thieno[2,3-d]pyrimidine-2,4-diamine (three-letter code: 63Y) (formula: C<sub>13</sub>H<sub>9</sub>F<sub>3</sub>N<sub>4</sub>S<sub>2</sub>).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	A	1	Total	C	F	N	S	0	0
			21	13	2	4	2		
3	A	1	Total	C	F	N	S	0	0
			21	13	2	4	2		

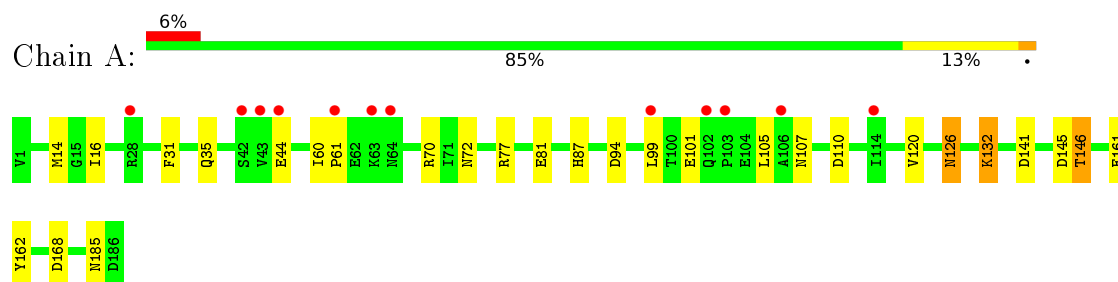
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	188	Total	O	0	0
			188	188		

### 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 ( $RSRZ > 2$ ). 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: Dihydrofolate reductase



## 4 Data and refinement statistics

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	84.61Å 84.61Å 78.09Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	26.72 – 1.21 26.72 – 1.21	Depositor EDS
% Data completeness (in resolution range)	98.7 (26.72-1.21) 98.7 (26.72-1.21)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	0.02	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.03 (at 1.21Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, $R_{free}$	0.193 , 0.219 0.190 , 0.215	Depositor DCC
$R_{free}$ test set	3157 reflections (5.35%)	DCC
Wilson B-factor (Å <sup>2</sup> )	13.7	Xtriage
Anisotropy	0.044	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.43 , 55.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.025 for h,-h-k,-l	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	1822	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	16.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.59% 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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: NDP, 63Y

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	1.35	7/1583 (0.4%)	1.25	11/2133 (0.5%)

All (7) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	126	ASN	CG-ND2	7.92	1.52	1.32
1	A	126	ASN	CB-CG	7.80	1.69	1.51
1	A	120	VAL	CB-CG1	-6.11	1.40	1.52
1	A	70	ARG	CZ-NH1	5.87	1.40	1.33
1	A	35	GLN	CG-CD	5.71	1.64	1.51
1	A	14	MET	SD-CE	-5.56	1.46	1.77
1	A	81	GLU	CD-OE1	-5.07	1.20	1.25

All (11) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	A	132	LYS	CD-CE-NZ	-9.66	89.47	111.70
1	A	168	ASP	CB-CG-OD1	9.20	126.58	118.30
1	A	70	ARG	NE-CZ-NH2	8.91	124.76	120.30
1	A	141	ASP	CB-CG-OD1	7.28	124.85	118.30
1	A	161	GLU	OE1-CD-OE2	-6.22	115.83	123.30
1	A	110	ASP	CB-CG-OD2	6.21	123.89	118.30
1	A	146	THR	OG1-CB-CG2	-6.20	95.74	110.00
1	A	70	ARG	NE-CZ-NH1	-6.12	117.24	120.30
1	A	77	ARG	NE-CZ-NH2	5.33	122.97	120.30
1	A	162	TYR	CB-CG-CD1	-5.27	117.84	121.00
1	A	94	ASP	CB-CG-OD2	-5.26	113.56	118.30

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	1544	0	1549	26	0
2	A	48	0	26	5	0
3	A	42	0	0	18	0
4	A	188	0	0	4	0
All	All	1822	0	1575	33	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 10.

All (33) 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:61:PRO:HD2	3:A:202:63Y:FAE	1.28	1.23
2:A:201:NDP:H42N	3:A:202:63Y:CAA	1.71	1.19
1:A:61:PRO:CD	3:A:202:63Y:FAE	1.78	1.19
1:A:99:LEU:HD22	1:A:105:LEU:HD12	1.45	0.98
1:A:99:LEU:CD2	1:A:105:LEU:HD12	1.94	0.96
1:A:31[A]:PHE:HE1	3:A:202:63Y:CAG	1.82	0.92
1:A:72:ASN:H	1:A:87:HIS:HD2	1.14	0.91
1:A:61:PRO:HD3	3:A:202:63Y:FAE	1.66	0.85
1:A:31[A]:PHE:CE1	3:A:202:63Y:CAN	2.64	0.79
1:A:126:ASN:OD1	4:A:301:HOH:O	2.03	0.76
1:A:72:ASN:H	1:A:87:HIS:CD2	2.03	0.75
2:A:201:NDP:C4N	3:A:202:63Y:CAA	2.59	0.75
1:A:31[A]:PHE:CE1	3:A:202:63Y:CAG	2.69	0.73
1:A:99:LEU:HD22	1:A:105:LEU:CD1	2.19	0.71
3:A:202:63Y:CAA	3:A:203:63Y:NAC	2.49	0.69
1:A:99:LEU:CD2	1:A:105:LEU:CD1	2.72	0.67
3:A:202:63Y:CAH	3:A:203:63Y:CAA	2.72	0.66
2:A:201:NDP:H42N	3:A:202:63Y:CAP	2.27	0.65
1:A:145:ASP:OD1	1:A:146:THR:HG23	1.95	0.65
1:A:31[A]:PHE:CZ	3:A:202:63Y:FAD	2.40	0.65
1:A:126:ASN:CG	4:A:301:HOH:O	2.44	0.56
1:A:31[A]:PHE:CE1	3:A:202:63Y:FAD	2.52	0.52
1:A:16:ILE:O	2:A:201:NDP:H2N	2.10	0.51

*Continued on next page...*



Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:61:PRO:HD2	3:A:202:63Y:CAR	2.37	0.50
1:A:132:LYS:HE2	4:A:340:HOH:O	2.12	0.48
1:A:132:LYS:CE	4:A:340:HOH:O	2.62	0.48
1:A:99:LEU:O	1:A:99:LEU:HD23	2.15	0.46
1:A:31[B]:PHE:HE1	3:A:203:63Y:SAK	2.39	0.46
2:A:201:NDP:C5N	3:A:202:63Y:CAA	2.94	0.45
1:A:107:ASN:N	1:A:107:ASN:OD1	2.39	0.44
3:A:202:63Y:CAH	3:A:203:63Y:CAP	2.67	0.43
1:A:60:ILE:HG23	3:A:202:63Y:CAN	2.50	0.41
1:A:99:LEU:HD23	1:A:105:LEU:HD12	1.92	0.41

There are no symmetry-related clashes.

## 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	190/186 (102%)	187 (98%)	3 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 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	174/168 (104%)	171 (98%)	3 (2%)	68	30

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	44	GLU
1	A	101	GLU
1	A	185	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	64	ASN
1	A	87	HIS
1	A	127	HIS
1	A	185	ASN

### 5.3.3 RNA [i](#)

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](#)

3 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	NDP	A	201	-	44,52,52	1.83	14 (31%)	55,80,80	2.26	10 (18%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
3	63Y	A	202	3	20,23,24	2.22	3 (15%)	20,34,36	1.72	2 (10%)
3	63Y	A	203	3	20,23,24	2.47	5 (25%)	20,34,36	2.08	6 (30%)

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
2	NDP	A	201	-	-	0/30/77/77	0/5/5/5
3	63Y	A	202	3	-	0/2/4/4	0/3/3/3
3	63Y	A	203	3	-	0/2/4/4	0/3/3/3

All (22) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	203	63Y	CAQ-SAK	-9.05	1.60	1.77
3	A	202	63Y	CAT-SAL	-6.90	1.60	1.72
3	A	202	63Y	CAP-C5	-4.98	1.35	1.40
3	A	202	63Y	C5-C4	-4.32	1.35	1.42
2	A	201	NDP	PN-O5D	-3.37	1.44	1.59
2	A	201	NDP	PA-O2A	-2.56	1.44	1.55
2	A	201	NDP	PA-O1A	-2.44	1.42	1.51
2	A	201	NDP	PN-O2N	-2.43	1.44	1.55
2	A	201	NDP	C2D-C1D	-2.42	1.45	1.53
2	A	201	NDP	P2B-O1X	-2.31	1.43	1.50
2	A	201	NDP	O4D-C4D	-2.10	1.40	1.45
3	A	203	63Y	CAA-CAP	2.12	1.56	1.51
3	A	203	63Y	CAS-CAR	2.18	1.41	1.37
2	A	201	NDP	C7N-N7N	2.27	1.39	1.33
2	A	201	NDP	C5D-C4D	2.80	1.60	1.51
3	A	203	63Y	CAG-CAH	2.93	1.44	1.38
2	A	201	NDP	C2A-N1A	3.22	1.40	1.33
2	A	201	NDP	C2N-N1N	3.32	1.43	1.37
3	A	203	63Y	CAT-SAL	3.35	1.78	1.72
2	A	201	NDP	O4B-C1B	3.50	1.46	1.41
2	A	201	NDP	C6N-C5N	3.52	1.39	1.33
2	A	201	NDP	C6N-N1N	3.74	1.47	1.37

All (18) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	201	NDP	N3A-C2A-N1A	-11.92	119.51	128.87
3	A	202	63Y	N3-C2-N1	-5.93	119.48	127.56
3	A	203	63Y	CAH-CAQ-CAS	-5.73	112.78	119.72
2	A	201	NDP	C1B-N9A-C4A	-3.91	122.44	126.81
2	A	201	NDP	O4D-C1D-N1N	-3.20	101.99	108.09
3	A	202	63Y	C5-C6-NAC	-2.91	117.00	122.73
2	A	201	NDP	C1D-N1N-C2N	-2.84	115.91	120.85
3	A	203	63Y	NAB-C2-N3	-2.52	112.99	117.72
2	A	201	NDP	O4B-C1B-C2B	-2.33	102.40	106.60
2	A	201	NDP	O4B-C4B-C3B	-2.18	100.74	105.16
3	A	203	63Y	N3-C2-N1	-2.10	124.70	127.56
3	A	203	63Y	FAD-CAN-CAG	2.19	123.44	118.46
2	A	201	NDP	C2D-C1D-N1N	2.61	120.38	113.42
2	A	201	NDP	O2B-P2B-O1X	2.83	114.23	107.48
3	A	203	63Y	NAB-C2-N1	2.97	122.10	117.20
2	A	201	NDP	C2A-N1A-C6A	3.38	124.79	118.77
3	A	203	63Y	CAG-CAH-CAQ	4.17	125.65	120.54
2	A	201	NDP	C4B-O4B-C1B	4.46	114.37	109.64

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 19 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	201	NDP	5	0
3	A	202	63Y	17	0
3	A	203	63Y	4	0

## 5.7 Other polymers ⓘ

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	186/186 (100%)	0.52	12 (6%)	22 19	9, 13, 33, 48	0

All (12) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	106	ALA	5.4
1	A	44	GLU	4.5
1	A	42	SER	4.0
1	A	63	LYS	3.7
1	A	103	PRO	3.1
1	A	102	GLN	2.9
1	A	28	ARG	2.9
1	A	43	VAL	2.8
1	A	61	PRO	2.3
1	A	114	ILE	2.2
1	A	64	ASN	2.2
1	A	99	LEU	2.0

### 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
3	63Y	A	202	21/22	0.92	0.17	2.00	11,14,27,28	21
3	63Y	A	203	21/22	0.95	0.11	0.33	7,11,16,21	21
2	NDP	A	201	48/48	0.96	0.10	0.32	10,14,18,19	0

## 6.5 Other polymers

There are no such residues in this entry.