



# Full wwPDB X-ray Structure Validation Report ⓘ

Sep 20, 2023 – 06:45 PM EDT

PDB ID : 5KI6  
Title : Human Argonaute-2 bound to a guide RNA with a nucleobase modification at position 1  
Authors : Schirle, N.T.; MacRae, I.J.  
Deposited on : 2016-06-16  
Resolution : 2.15 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

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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 : 2.35.1  
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) : 2.35.1

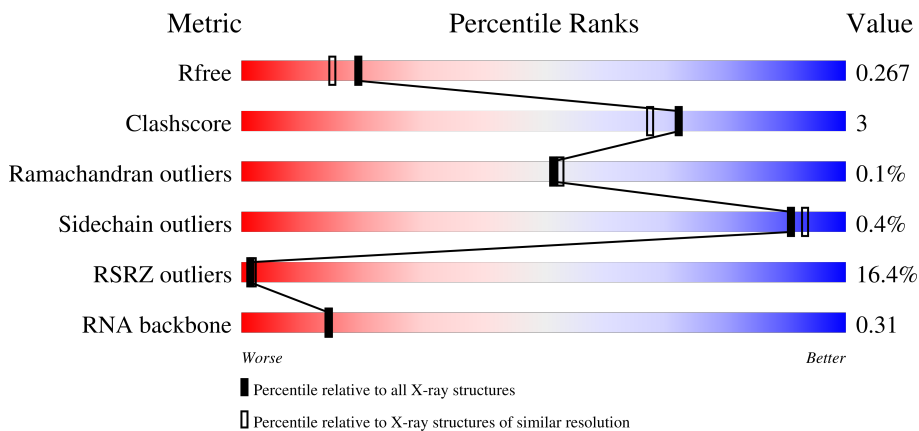
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.15 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	1479 (2.16-2.16)
Clashscore	141614	1585 (2.16-2.16)
Ramachandran outliers	138981	1560 (2.16-2.16)
Sidechain outliers	138945	1559 (2.16-2.16)
RSRZ outliers	127900	1456 (2.16-2.16)
RNA backbone	3102	1060 (2.60-1.72)

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 of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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	859	 15% 85% 8% 7%
2	B	21	 19% 19% 10% 52%

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 6756 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 Protein argonaute-2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	798	6406	4078	1154	1135	39	0	0	0

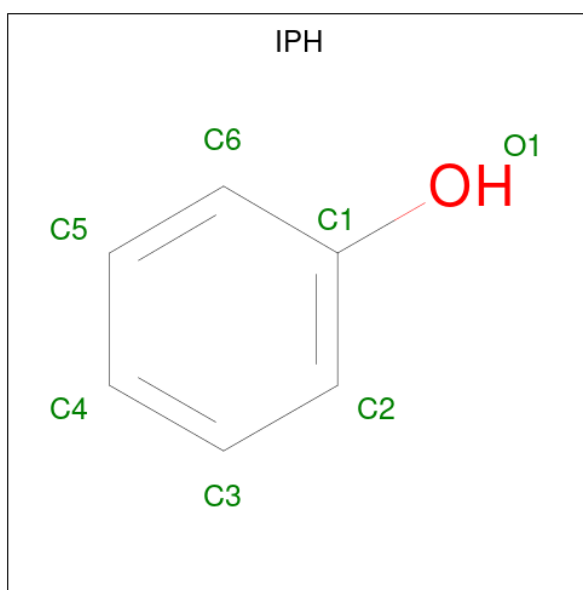
There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	387	ASP	SER	engineered mutation	UNP Q9UKV8

- Molecule 2 is a RNA chain called miR-122.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	B	10	224	103	38	73	10	0	0	0

- Molecule 3 is PHENOL (three-letter code: IPH) (formula: C<sub>6</sub>H<sub>6</sub>O).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 7 6 1	0	0
3	A	1	Total C O 7 6 1	0	0
3	A	1	Total C O 7 6 1	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	100	Total O 100 100	0	0
4	B	5	Total O 5 5	0	0



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	63.28Å 107.30Å 68.86Å 90.00° 107.44° 90.00°	Depositor
Resolution (Å)	47.57 – 2.15 47.57 – 2.15	Depositor EDS
% Data completeness (in resolution range)	99.2 (47.57-2.15) 99.3 (47.57-2.15)	Depositor EDS
$R_{merge}$	0.04	Depositor
$R_{sym}$	0.03	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.17 (at 2.16Å)	Xtrriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, $R_{free}$	0.237 , 0.270 0.236 , 0.267	Depositor DCC
$R_{free}$ test set	2382 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	46.7	Xtrriage
Anisotropy	0.390	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 47.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	6756	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	70.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.47% 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: IPH, 6U0

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	0.30	0/6554	0.44	2/8867 (0.0%)
2	B	1.19	3/215 (1.4%)	1.22	6/331 (1.8%)
All	All	0.37	3/6769 (0.0%)	0.49	8/9198 (0.1%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	6	U	O3'-P	-6.90	1.52	1.61
2	B	5	G	O3'-P	-6.52	1.53	1.61
2	B	4	A	O3'-P	-5.83	1.54	1.61

All (8) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	6	U	C1'-C2'-O2'	-7.36	88.53	110.60
2	B	4	A	C1'-C2'-O2'	-6.36	91.53	110.60
2	B	6	U	N1-C1'-C2'	-5.78	105.65	112.00
2	B	5	G	N9-C1'-C2'	-5.59	105.85	112.00
1	A	761	ARG	C-N-CD	5.58	140.13	128.40
1	A	108	ILE	N-CA-C	-5.41	96.40	111.00
2	B	5	G	C1'-C2'-O2'	-5.36	94.51	110.60
2	B	4	A	N9-C1'-C2'	-5.08	106.41	112.00

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	6406	0	6467	43	0
2	B	224	0	97	4	0
3	A	21	0	18	0	0
4	A	100	0	0	0	0
4	B	5	0	0	0	0
All	All	6756	0	6582	45	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 3.

All (45) 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:190:ASN:HD22	1:A:198:VAL:HB	1.24	0.98
1:A:190:ASN:ND2	1:A:198:VAL:HB	1.82	0.95
1:A:192:LEU:HD11	1:A:198:VAL:CG2	2.21	0.70
1:A:192:LEU:HD11	1:A:198:VAL:HG23	1.74	0.68
1:A:192:LEU:HD22	1:A:360:GLN:O	2.00	0.62
1:A:541:GLY:HA3	1:A:851:ASP:HB2	1.87	0.56
1:A:606:GLY:O	1:A:608:LYS:N	2.40	0.55
1:A:190:ASN:N	1:A:191:PRO:CD	2.68	0.55
1:A:108:ILE:O	1:A:108:ILE:HG12	2.07	0.53
1:A:502:GLU:O	1:A:506:ARG:N	2.40	0.52
1:A:566:LYS:NZ	1:A:859:ALA:OXT	2.41	0.50
1:A:59:LEU:N	1:A:98:LYS:O	2.42	0.50
1:A:365:ILE:HD11	2:B:7:G:C5	2.48	0.49
1:A:332:GLN:O	1:A:334:GLN:N	2.43	0.49
1:A:174:TYR:HD1	1:A:183:THR:HG23	1.79	0.48
1:A:192:LEU:HD23	1:A:363:THR:HB	1.96	0.48
1:A:22:ALA:HA	1:A:684:LEU:HD23	1.96	0.47
1:A:526:THR:O	1:A:528:VAL:N	2.42	0.47
1:A:606:GLY:C	1:A:608:LYS:H	2.18	0.47
1:A:193:GLY:O	1:A:360:GLN:NE2	2.45	0.46
1:A:848:VAL:HG22	1:A:852:THR:HB	1.98	0.46
1:A:583:ARG:NH1	1:A:620:ALA:O	2.49	0.46

*Continued on next page...*



Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:505:PHE:HB3	1:A:542:MET:HE1	1.98	0.46
1:A:651:ILE:HD11	1:A:694:LEU:HD11	1.99	0.45
1:A:332:GLN:O	1:A:333:GLU:HB2	2.17	0.44
1:A:177:VAL:HG11	1:A:351:ARG:HD3	1.99	0.43
1:A:511:THR:HG23	1:A:512:TYR:CD2	2.53	0.43
1:A:657:THR:O	1:A:658:ARG:HG2	2.18	0.43
1:A:169:LEU:HB3	1:A:170:PRO:HD3	1.98	0.43
2:B:1:6U0:O2'	2:B:2:G:OP1	2.35	0.43
1:A:147:LEU:HD21	1:A:213:MET:CE	2.48	0.43
1:A:54:ILE:HB	1:A:138:VAL:HB	2.01	0.42
2:B:5:G:C6	2:B:6:U:O4	2.73	0.42
1:A:112:LYS:HG3	1:A:113:VAL:N	2.34	0.42
1:A:606:GLY:C	1:A:608:LYS:N	2.73	0.42
1:A:757:GLN:OE1	2:B:5:G:N2	2.50	0.42
1:A:801:ALA:N	1:A:802:PRO:HD2	2.34	0.42
1:A:192:LEU:CD1	1:A:198:VAL:HG23	2.45	0.42
1:A:498:ALA:HB1	1:A:531:GLU:HG3	2.01	0.42
1:A:202:PHE:CB	1:A:382:LEU:HD21	2.50	0.42
1:A:372:ALA:HB3	1:A:373:PRO:HD3	2.03	0.41
1:A:151:LEU:HB3	1:A:744:THR:HG21	2.03	0.40
1:A:443:HIS:ND1	1:A:444:THR:HG23	2.36	0.40
1:A:108:ILE:O	1:A:108:ILE:CG1	2.69	0.40
1:A:412:VAL:HG22	1:A:734:THR:HG22	2.03	0.40

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	784/859 (91%)	755 (96%)	28 (4%)	1 (0%)	51   53

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	607	LYS

### 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	706/752 (94%)	703 (100%)	3 (0%)	<a href="#">91</a> <a href="#">93</a>

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

Mol	Chain	Res	Type
1	A	23	PHE
1	A	108	ILE
1	A	804	TYR

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

Mol	Chain	Res	Type
1	A	190	ASN
1	A	216	ASN

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	B	7/21 (33%)	0	0

There are no RNA backbone outliers to report.

There are no RNA pucker outliers to report.

## 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 monosaccharides 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	$\# Z  > 2$
3	IPH	A	901	-	7,7,7	0.69	0	8,8,8	0.25	0
3	IPH	A	902	-	7,7,7	0.69	0	8,8,8	0.25	0
3	IPH	A	903	-	7,7,7	0.71	0	8,8,8	0.25	0

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	IPH	A	901	-	-	-	0/1/1/1
3	IPH	A	902	-	-	-	0/1/1/1
3	IPH	A	903	-	-	-	0/1/1/1

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.

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

### 6.1 Protein, DNA and RNA chains

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	798/859 (92%)	1.01	132 (16%) <b>1</b> <b>2</b>	30, 66, 119, 174	0
2	B	9/21 (42%)	-0.23	0 <b>100</b> <b>100</b>	45, 55, 87, 112	0
All	All	807/880 (91%)	1.00	132 (16%) <b>1</b> <b>2</b>	30, 66, 119, 174	0

All (132) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	270	THR	6.3
1	A	353	ILE	6.2
1	A	505	PHE	6.2
1	A	151	LEU	5.9
1	A	86	ILE	5.5
1	A	193	GLY	5.3
1	A	110	ARG	4.9
1	A	517	LEU	4.9
1	A	552	VAL	4.8
1	A	540	LEU	4.7
1	A	196	ARG	4.7
1	A	85	GLN	4.6
1	A	332	GLN	4.2
1	A	448	ILE	4.2
1	A	277	ARG	4.1
1	A	604	GLY	4.1
1	A	77	HIS	4.1
1	A	244	GLU	4.1
1	A	510	ASN	4.0
1	A	347	VAL	4.0
1	A	449	LYS	3.9
1	A	195	GLY	3.9
1	A	526	THR	3.8
1	A	184	ALA	3.8

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Res</b>	<b>Type</b>	<b>RSRZ</b>
1	A	194	GLY	3.8
1	A	156	PHE	3.7
1	A	240	PHE	3.7
1	A	243	ILE	3.7
1	A	335	LYS	3.7
1	A	423	ARG	3.7
1	A	854	ARG	3.6
1	A	211	TRP	3.6
1	A	571	LEU	3.5
1	A	485	ILE	3.5
1	A	355	LYS	3.5
1	A	60	ASP	3.4
1	A	80	GLN	3.3
1	A	356	LEU	3.3
1	A	496	GLN	3.3
1	A	117	VAL	3.3
1	A	173	ARG	3.3
1	A	133	LYS	3.3
1	A	501	VAL	3.3
1	A	108	ILE	3.2
1	A	148	SER	3.2
1	A	451	TRP	3.2
1	A	512	TYR	3.2
1	A	111	ASP	3.2
1	A	90	ARG	3.2
1	A	361	THR	3.2
1	A	157	GLU	3.2
1	A	134	TRP	3.2
1	A	605	ASP	3.1
1	A	190	ASN	3.1
1	A	358	ASP	3.1
1	A	352	CYS	3.1
1	A	534	ARG	3.1
1	A	607	LYS	3.1
1	A	126	ARG	3.1
1	A	470	PHE	3.1
1	A	336	HIS	3.0
1	A	513	ALA	3.0
1	A	210	LEU	2.9
1	A	477	ILE	2.9
1	A	260	LYS	2.9
1	A	135	VAL	2.9

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Res</b>	<b>Type</b>	<b>RSRZ</b>
1	A	149	GLY	2.9
1	A	511	THR	2.9
1	A	783	THR	2.9
1	A	787	CYS	2.8
1	A	486	GLN	2.8
1	A	522	LEU	2.8
1	A	658	ARG	2.8
1	A	241	LYS	2.8
1	A	127	ILE	2.7
1	A	59	LEU	2.7
1	A	87	PHE	2.7
1	A	132	ILE	2.6
1	A	81	HIS	2.6
1	A	547	VAL	2.6
1	A	635	ARG	2.6
1	A	773	ARG	2.6
1	A	425	LYS	2.6
1	A	253	SER	2.5
1	A	360	GLN	2.5
1	A	544	THR	2.5
1	A	460	ARG	2.5
1	A	671	VAL	2.5
1	A	106	LEU	2.4
1	A	491	PHE	2.4
1	A	304	VAL	2.4
1	A	199	TRP	2.4
1	A	841	ALA	2.4
1	A	130	VAL	2.4
1	A	22	ALA	2.3
1	A	64	GLU	2.3
1	A	198	VAL	2.3
1	A	394	VAL	2.3
1	A	546	CYS	2.3
1	A	82	PHE	2.3
1	A	256	VAL	2.3
1	A	366	ARG	2.3
1	A	493	LYS	2.3
1	A	250	LEU	2.3
1	A	694	LEU	2.2
1	A	438	ARG	2.2
1	A	479	ARG	2.2
1	A	116	GLU	2.2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	439	ASN	2.2
1	A	851	ASP	2.2
1	A	57	TYR	2.2
1	A	224	PHE	2.2
1	A	351	ARG	2.2
1	A	225	TYR	2.2
1	A	322	TYR	2.2
1	A	450	VAL	2.1
1	A	227	ALA	2.1
1	A	330	VAL	2.1
1	A	349	GLY	2.1
1	A	136	SER	2.1
1	A	131	SER	2.1
1	A	530	ALA	2.1
1	A	734	THR	2.1
1	A	815	TYR	2.1
1	A	455	CYS	2.1
1	A	84	THR	2.0
1	A	539	VAL	2.0
1	A	279	TYR	2.0
1	A	266	LYS	2.0
1	A	839	HIS	2.0
1	A	238	LEU	2.0
1	A	154	VAL	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 monosaccharides 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, 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	B-factors( $\text{\AA}^2$ )	Q<0.9
3	IPH	A	903	7/7	0.80	0.18	57,59,63,72	0
3	IPH	A	902	7/7	0.90	0.17	66,66,69,71	0
3	IPH	A	901	7/7	0.92	0.17	46,54,60,62	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.