



# Full wwPDB X-ray Structure Validation Report ⓘ

Mar 8, 2018 – 11:49 pm GMT

PDB ID : 5KPQ  
Title : Structure of human PARP1 catalytic domain bound to a quinazoline-2,4(1H,3H)-dione inhibitor  
Authors : Cao, R.; Wang, Y.L.; Zhou, J.; Huang, N.; Xu, B.L.  
Deposited on : 2016-07-05  
Resolution : 2.55 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.7.3 (157068), CSD as539be (2018)  
Xtriage (Phenix) : 1.13  
EDS : trunk30967  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Refmac : 5.8.0158  
CCP4 : 7.0 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk30967

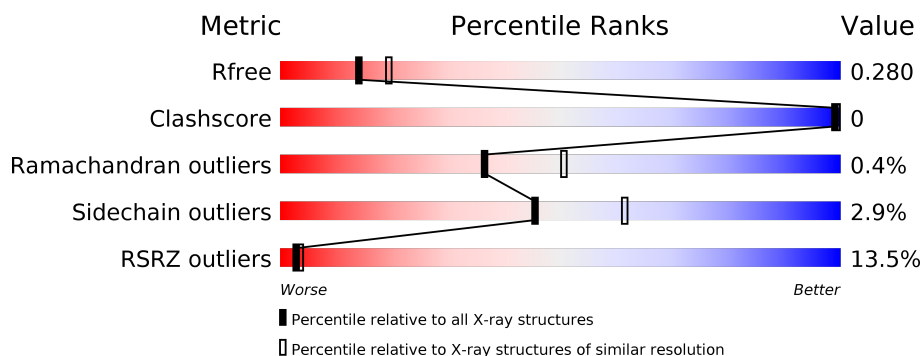
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.55 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))
$R_{free}$	111664	1053 (2.56-2.52)
Clashscore	122126	1098 (2.56-2.52)
Ramachandran outliers	120053	1088 (2.56-2.52)
Sidechain outliers	120020	1088 (2.56-2.52)
RSRZ outliers	108989	1043 (2.56-2.52)

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	350	<div> <div>9%</div> <div>97%</div> <div>.</div> </div>
1	B	350	<div> <div>18%</div> <div>93%</div> <div>5% .</div> </div>

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 5633 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.

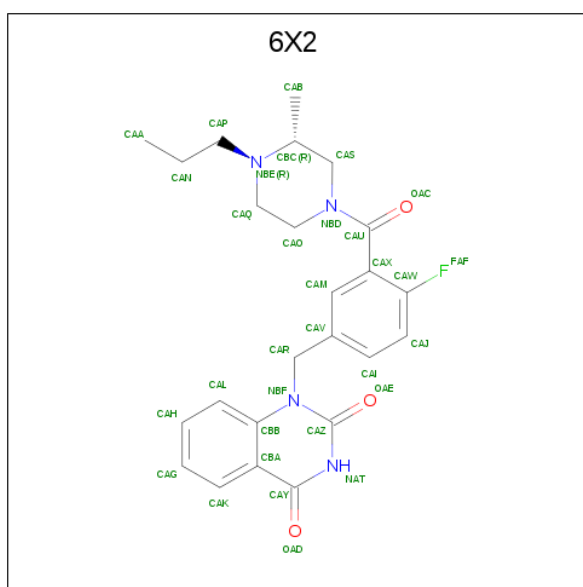
- Molecule 1 is a protein called Poly [ADP-ribose] polymerase 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	349	Total	C	N	O	S	0	3	0
			2755	1752	466	525	12			
1	B	349	Total	C	N	O	S	0	3	0
			2755	1752	466	525	12			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	762	ALA	VAL	engineered mutation	UNP P09874
B	762	ALA	VAL	engineered mutation	UNP P09874

- Molecule 2 is 1-[[4-fluoranyl-3-[(3R)-3-methyl-4-propyl-piperazin-1-yl]carbonyl-phenyl]methyl]quinazoline-2,4-dione (three-letter code: 6X2) (formula: C<sub>24</sub>H<sub>27</sub>FN<sub>4</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	F	N	O	0	0
			32	24	1	4	3		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	B	1	Total	C	F	N	O	0	0
			32	24	1	4	3		

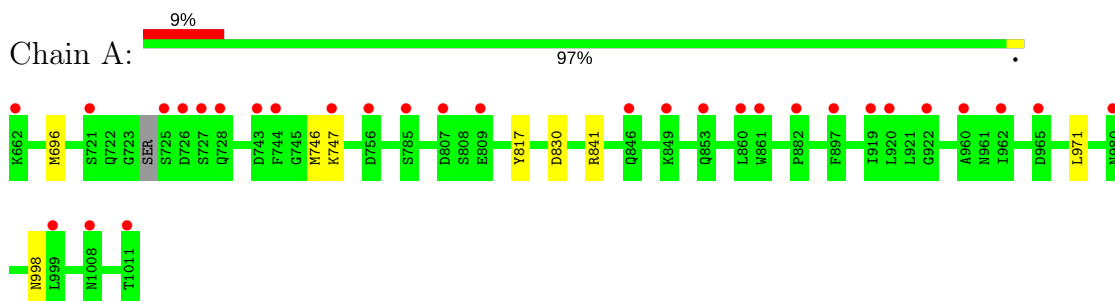
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	35	Total	O	0	0
			35	35		
3	B	24	Total	O	0	0
			24	24		

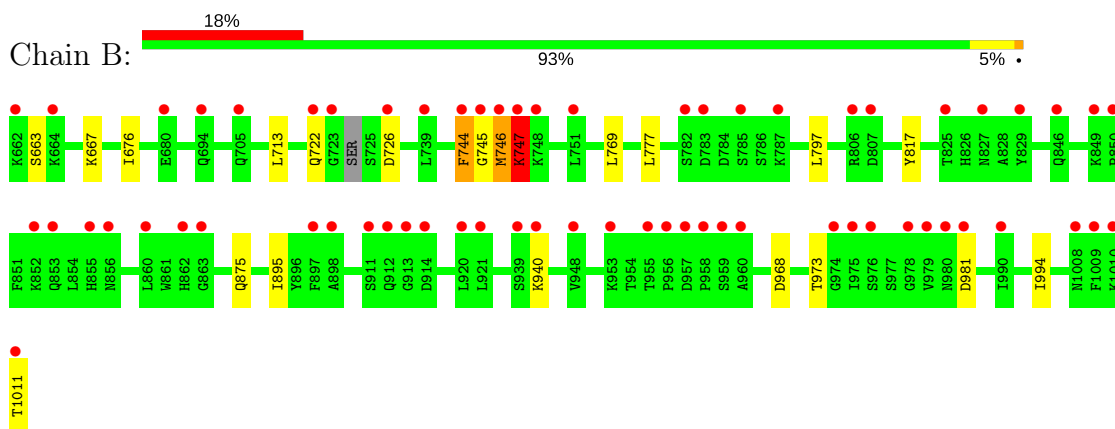
### 3 Residue-property plots [i](#)

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: Poly [ADP-ribose] polymerase 1



- Molecule 1: Poly [ADP-ribose] polymerase 1



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	48.63Å 92.47Å 163.85Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	81.92 – 2.55 27.31 – 2.55	Depositor EDS
% Data completeness (in resolution range)	99.2 (81.92-2.55) 99.3 (27.31-2.55)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.72 (at 2.54Å)	Xtriage
Refinement program	REFMAC 5.8.0123	Depositor
R, $R_{free}$	0.233 , 0.282 0.234 , 0.280	Depositor DCC
$R_{free}$ test set	1199 reflections (4.86%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	52.4	Xtriage
Anisotropy	0.576	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 40.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	5633	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	58.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.33% 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: 6X2

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	0.36	0/2818	0.60	1/3802 (0.0%)
1	B	0.37	0/2818	0.58	0/3802
All	All	0.37	0/5636	0.59	1/7604 (0.0%)

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	2
1	B	0	3
All	All	0	5

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	841	ARG	NE-CZ-NH1	5.61	123.10	120.30

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	746	MET	Peptide
1	A	747	LYS	Peptide
1	B	745	GLY	Peptide
1	B	746	MET	Peptide
1	B	747	LYS	Peptide

## 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	2755	0	2790	0	0
1	B	2755	0	2790	4	0
2	A	32	0	0	0	0
2	B	32	0	0	0	0
3	A	35	0	0	0	0
3	B	24	0	0	0	0
All	All	5633	0	5580	4	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 0.

All (4) 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:B:895:ILE:HD11	1:B:994:ILE:HG22	1.82	0.60
1:B:744:PHE:HB2	1:B:747:LYS:HA	1.90	0.54
1:B:746:MET:C	1:B:747:LYS:HG2	2.39	0.43
1:B:676:ILE:HD11	1:B:797:LEU:HD11	2.03	0.41

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	348/350 (99%)	343 (99%)	5 (1%)	0	<b>100</b> <b>100</b>

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	348/350 (99%)	329 (94%)	16 (5%)	3 (1%)	19	28
All	All	696/700 (99%)	672 (97%)	21 (3%)	3 (0%)	36	50

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	722	GLN
1	B	663	SER
1	B	747	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	309/307 (101%)	304 (98%)	5 (2%)	65	80
1	B	309/307 (101%)	296 (96%)	13 (4%)	32	45
All	All	618/614 (101%)	600 (97%)	18 (3%)	45	61

All (18) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	696	MET
1	A	817	TYR
1	A	830	ASP
1	A	971	LEU
1	A	998	ASN
1	B	667	LYS
1	B	713	LEU
1	B	726	ASP
1	B	744	PHE
1	B	769	LEU
1	B	777	LEU
1	B	817	TYR
1	B	875	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	940	LYS
1	B	968	ASP
1	B	973	THR
1	B	981	ASP
1	B	1011	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	961	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](#)

2 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	6X2	A	1101	-	32,35,35	1.74	5 (15%)	40,50,50	2.14	9 (22%)
2	6X2	B	1101	-	32,35,35	1.78	4 (12%)	40,50,50	2.28	6 (15%)

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	6X2	A	1101	-	-	0/15/28/28	1/4/4/4
2	6X2	B	1101	-	-	0/15/28/28	1/4/4/4

All (9) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	1101	6X2	CAX-CAU	-6.77	1.39	1.50
2	A	1101	6X2	CAX-CAU	-6.16	1.40	1.50
2	B	1101	6X2	CAR-CAV	-4.56	1.41	1.51
2	A	1101	6X2	CAR-CAV	-4.38	1.41	1.51
2	A	1101	6X2	CBB-NBF	-2.55	1.37	1.40
2	A	1101	6X2	FAF-CAW	-2.17	1.30	1.35
2	B	1101	6X2	CAS-NBD	2.21	1.49	1.46
2	A	1101	6X2	CAY-NAT	3.18	1.38	1.33
2	B	1101	6X2	CAY-NAT	3.53	1.39	1.33

All (15) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	1101	6X2	CBA-CAY-NAT	-8.48	118.58	124.45
2	A	1101	6X2	CBA-CAY-NAT	-8.26	118.74	124.45
2	B	1101	6X2	CAJ-CAW-CAX	-2.99	119.68	123.11
2	A	1101	6X2	CAN-CAP-NBE	-2.89	108.70	114.87
2	A	1101	6X2	CAJ-CAW-CAX	-2.22	120.56	123.11
2	A	1101	6X2	CAW-CAX-CAU	-2.13	119.89	124.88
2	B	1101	6X2	CAL-CBB-CBA	-2.04	117.17	120.02
2	A	1101	6X2	CAQ-NBE-CBC	2.20	115.45	110.49
2	B	1101	6X2	CAM-CAX-CAW	2.44	119.85	116.82
2	A	1101	6X2	CAR-NBF-CBB	2.47	123.61	120.15
2	A	1101	6X2	CAO-NBD-CAS	2.61	118.53	113.10
2	A	1101	6X2	CAK-CBA-CBB	3.96	119.86	118.04
2	A	1101	6X2	CAY-NAT-CAZ	5.94	120.20	115.14
2	B	1101	6X2	CAY-NAT-CAZ	6.47	120.65	115.14
2	B	1101	6X2	CAK-CBA-CBB	6.55	121.06	118.04

There are no chirality outliers.

There are no torsion outliers.

All (2) ring outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	1101	6X2	CAO-CAQ-CAS-CBC-NBD-NBE
2	B	1101	6X2	CAO-CAQ-CAS-CBC-NBD-NBE

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 [i](#)

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	349/350 (99%)	0.47	30 (8%) 10 12	30, 49, 74, 95	0
1	B	349/350 (99%)	1.11	64 (18%) 1 1	38, 62, 97, 153	0
All	All	698/700 (99%)	0.79	94 (13%) 3 3	30, 55, 93, 153	0

All (94) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	747	LYS	11.3
1	B	746	MET	10.7
1	B	1011	THR	8.7
1	B	748	LYS	6.1
1	A	807	ASP	6.1
1	B	975	ILE	6.0
1	B	662	LYS	5.7
1	B	745	GLY	5.6
1	B	974	GLY	5.4
1	A	662	LYS	5.2
1	B	853	GLN	5.0
1	B	959	SER	5.0
1	B	958	PRO	4.9
1	B	744	PHE	4.8
1	B	978	GLY	4.8
1	B	723	GLY	4.8
1	B	807	ASP	4.5
1	B	979	VAL	4.5
1	B	976	SER	4.3
1	A	727	SER	4.1
1	B	846	GLN	4.1
1	A	846	GLN	4.0
1	A	747	LYS	3.9
1	B	849	LYS	3.8

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	B	825	THR	3.8
1	B	920	LEU	3.8
1	B	726	ASP	3.6
1	A	965	ASP	3.6
1	A	726	ASP	3.5
1	B	897	PHE	3.5
1	B	827	ASN	3.5
1	B	1009	PHE	3.5
1	B	960	ALA	3.5
1	A	728	GLN	3.5
1	A	980	ASN	3.4
1	B	939	SER	3.4
1	B	705	GLN	3.3
1	B	980	ASN	3.3
1	B	860	LEU	3.3
1	B	921	LEU	3.3
1	B	664	LYS	3.2
1	B	850	PRO	3.2
1	B	956	PRO	3.2
1	B	913	GLY	3.1
1	A	785	SER	3.1
1	B	898	ALA	3.1
1	B	783	ASP	3.1
1	B	981	ASP	3.1
1	A	725	SER	3.1
1	B	785	SER	3.0
1	A	861	TRP	3.0
1	B	912	GLN	2.9
1	B	787	LYS	2.9
1	B	722	GLN	2.9
1	B	914	ASP	2.8
1	B	852	LYS	2.8
1	B	1008[A]	ASN	2.8
1	A	920	LEU	2.7
1	B	855	HIS	2.7
1	A	809	GLU	2.6
1	B	739	LEU	2.6
1	A	882	PRO	2.6
1	A	1011	THR	2.5
1	B	955	THR	2.5
1	A	922	GLY	2.5
1	A	1008[A]	ASN	2.5

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	743	ASP	2.5
1	B	863	GLY	2.5
1	A	962	ILE	2.4
1	A	919	ILE	2.4
1	B	856[A]	ASN	2.4
1	B	694	GLN	2.4
1	A	860	LEU	2.4
1	B	782	SER	2.3
1	A	960	ALA	2.3
1	B	911	SER	2.3
1	A	853	GLN	2.3
1	B	751	LEU	2.3
1	B	862	HIS	2.3
1	A	999	LEU	2.2
1	A	721	SER	2.2
1	B	680	GLU	2.2
1	A	744	PHE	2.2
1	B	953	LYS	2.2
1	B	990	ILE	2.2
1	B	806	ARG	2.2
1	A	756	ASP	2.1
1	B	957	ASP	2.1
1	B	948	VAL	2.1
1	A	897	PHE	2.1
1	B	1010	LYS	2.0
1	B	829	TYR	2.0
1	A	849	LYS	2.0
1	B	940	LYS	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. 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
2	6X2	B	1101	32/32	0.93	0.20	48,53,62,66	0
2	6X2	A	1101	32/32	0.94	0.18	32,40,51,53	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.