



# Full wwPDB X-ray Structure Validation Report ⓘ

Apr 19, 2026 – 12:39 AM UTC

PDB ID : 5PY8 / pdb\_00005py8  
Title : PanDDA analysis group deposition – Crystal Structure of SP100 after initial refinement with no ligand modelled (structure 68)  
Authors : Pearce, N.M.; Krojer, T.; Talon, R.; Bradley, A.R.; Fairhead, M.; Sethi, R.; Wright, N.; MacLean, E.; Collins, P.; Brandao-Neto, J.; Douangamath, A.; Renjie, Z.; Dias, A.; Ng, J.; Brennan, P.E.; Cox, O.; Bountra, C.; Arrowsmith, C.H.; Edwards, A.; von Delft, F.  
Deposited on : 2017-02-08  
Resolution : 1.66 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 : 4-5-2 with Phenix2.0  
Mogul : 2022.3.0, CSD as543be (2022)  
Xtriage (Phenix) : 2.0  
EDS : 3.0  
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)  
CCP4 : 9.0.010 (Gargrove)  
Density-Fitness : 1.0.12  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)

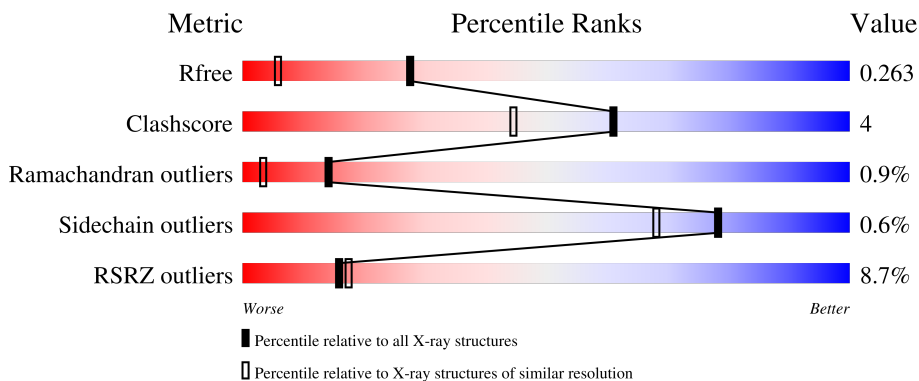
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.66 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))
$R_{free}$	180053	2563 (1.66-1.66)
Clashscore	190562	2662 (1.66-1.66)
Ramachandran outliers	187476	2621 (1.66-1.66)
Sidechain outliers	187428	2621 (1.66-1.66)
RSRZ outliers	180081	2564 (1.66-1.66)

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	180	
1	B	180	

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 3620 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.

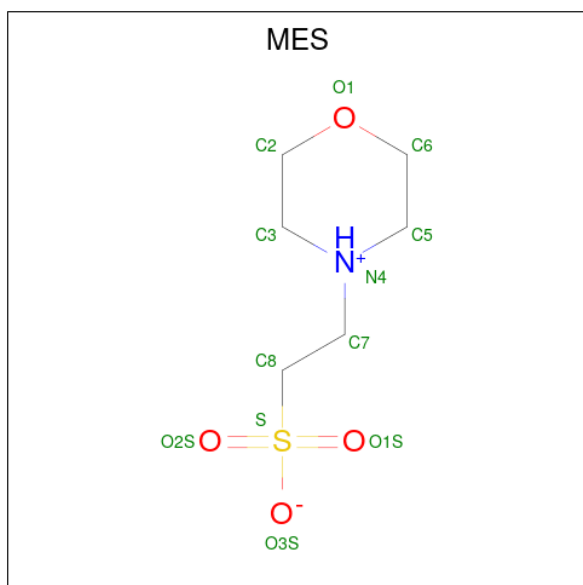
- Molecule 1 is a protein called Nuclear autoantigen Sp-100.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	178	Total 1536	C 971	N 268	O 278	S 19	0	10	0
1	B	177	Total 1480	C 940	N 258	O 263	S 19	0	7	0

- Molecule 2 is ZINC ION (CCD ID: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	3	Total 3	Zn 3	0	0
2	B	2	Total 2	Zn 2	0	0

- Molecule 3 is 2-(N-MORPHOLINO)-ETHANESULFONIC ACID (CCD ID: MES) (formula: C<sub>6</sub>H<sub>13</sub>NO<sub>4</sub>S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	S		
3	A	1	12	6	1	4	1	0	0

- Molecule 4 is 1,2-ETHANEDIOL (CCD ID: EDO) (formula: C<sub>2</sub>H<sub>6</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
4	A	1	4	2	2	0	0
4	B	1	4	2	2	0	0


- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	316	316	316	0	3
5	B	263	263	263	0	0

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


These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

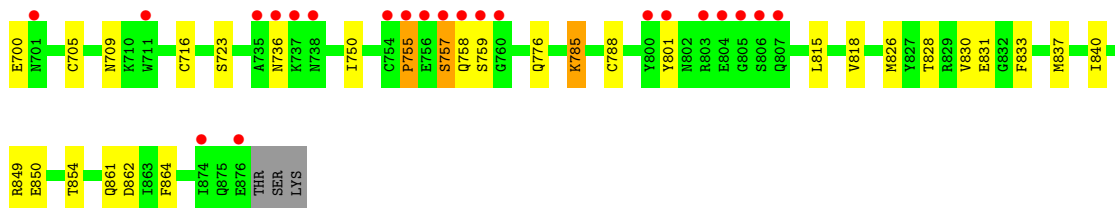
- Molecule 1: Nuclear autoantigen Sp-100

Chain A: 



- Molecule 1: Nuclear autoantigen Sp-100

Chain B: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	127.71Å 45.30Å 83.38Å 90.00° 102.26° 90.00°	Depositor
Resolution (Å)	38.91 – 1.66 38.91 – 1.66	Depositor EDS
% Data completeness (in resolution range)	95.1 (38.91-1.66) 95.1 (38.91-1.66)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.06 (at 1.66Å)	Xtrriage
Refinement program	REFMAC 5.8.0135	Depositor
R, $R_{free}$	0.207 , 0.252 0.219 , 0.263	Depositor DCC
$R_{free}$ test set	2768 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	22.5	Xtrriage
Anisotropy	0.103	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 45.9	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	0.95	EDS
Total number of atoms	3620	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	32.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.25% 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: MES, ZN, EDO

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	1.36	6/1603 (0.4%)	1.21	2/2154 (0.1%)
1	B	1.44	14/1538 (0.9%)	1.15	3/2071 (0.1%)
All	All	1.40	20/3141 (0.6%)	1.18	5/4225 (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
1	B	0	1

All (20) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	801	TYR	CA-C	8.80	1.56	1.52
1	B	830	VAL	N-CA	8.32	1.56	1.46
1	A	783	LEU	N-CA	-7.57	1.37	1.46
1	A	776	GLN	C-O	7.11	1.32	1.24
1	B	837	MET	N-CA	7.08	1.54	1.46
1	B	864	PHE	N-CA	6.89	1.54	1.46
1	A	780	GLU	C-O	6.43	1.31	1.24
1	B	788	CYS	N-CA	6.13	1.54	1.46
1	B	861	GLN	N-CA	5.84	1.53	1.46
1	B	785	LYS	N-CA	5.72	1.53	1.46
1	B	705	CYS	N-CA	5.36	1.52	1.46
1	B	818	VAL	N-CA	5.35	1.52	1.46
1	A	864	PHE	C-O	-5.28	1.18	1.24
1	B	840	ILE	CA-C	5.13	1.59	1.52
1	A	708	CYS	N-CA	5.13	1.52	1.46
1	A	860	VAL	C-O	-5.11	1.18	1.24

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	833	PHE	N-CA	5.11	1.52	1.46
1	B	815	LEU	C-O	-5.10	1.18	1.24
1	B	828	THR	CA-C	-5.08	1.45	1.52
1	B	723	SER	C-O	5.02	1.30	1.23

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	794	PHE	N-CA-C	6.55	120.37	112.38
1	A	849	ARG	CB-CA-C	-6.35	108.59	117.23
1	B	776	GLN	N-CA-C	-5.68	104.99	111.07
1	B	736	ASN	N-CA-C	5.58	118.12	111.71
1	B	831	GLU	N-CA-C	5.03	119.68	111.37

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	755	PRO	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	1536	0	1496	11	0
1	B	1480	0	1414	11	0
2	A	3	0	0	0	0
2	B	2	0	0	0	0
3	A	12	0	13	2	0
4	A	4	0	6	1	0
4	B	4	0	6	1	0
5	A	316	0	0	9	0
5	B	263	0	0	8	1
All	All	3620	0	2935	25	1

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 4.

All (25) 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)
4:B:903:EDO:O2	5:B:1001:HOH:O	1.73	1.06
1:A:829[A]:ARG:NH1	5:A:1001:HOH:O	1.93	0.99
1:A:803[A]:ARG:NH2	5:A:1002:HOH:O	1.97	0.97
3:A:904:MES:H81	5:A:1220:HOH:O	1.71	0.89
1:A:803[B]:ARG:NH1	5:A:1004:HOH:O	2.20	0.75
1:B:716[A]:CYS:SG	5:B:1175:HOH:O	2.45	0.75
1:A:831:GLU:OE2	5:A:1003:HOH:O	2.04	0.74
1:B:826[B]:MET:SD	5:B:1024:HOH:O	2.52	0.68
1:A:824:GLU:OE1	1:A:826[B]:MET:HE2	1.96	0.64
1:A:803[A]:ARG:NE	3:A:904:MES:O1S	2.34	0.57
1:A:734:GLU:OE1	1:A:740:TRP:NE1	2.36	0.54
1:B:826[B]:MET:HE2	5:B:1203:HOH:O	2.11	0.50
1:B:849:ARG:O	1:B:854[B]:THR:HG21	2.12	0.49
1:B:700:GLU:N	5:B:1005:HOH:O	2.44	0.49
1:B:757:SER:O	1:B:759:SER:N	2.42	0.47
1:B:757:SER:C	1:B:759:SER:N	2.73	0.46
4:A:905:EDO:C2	5:A:1106:HOH:O	2.63	0.45
1:B:709:ASN:ND2	5:B:1007:HOH:O	2.47	0.43
1:A:849:ARG:O	1:A:854[B]:THR:HG21	2.19	0.43
1:B:750:ILE:HD11	5:B:1001:HOH:O	2.19	0.42
1:B:757:SER:C	1:B:759:SER:H	2.26	0.42
1:B:785:LYS:HE3	5:B:1234:HOH:O	2.20	0.42
1:A:756:GLU:HG3	5:A:1142:HOH:O	2.19	0.42
1:A:762:GLN:NE2	5:A:1022:HOH:O	2.52	0.42
1:A:803[B]:ARG:HG3	5:A:1220:HOH:O	2.22	0.40

All (1) 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)
5:B:1252:HOH:O	5:B:1252:HOH:O[2_555]	0.91	1.29

## 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	186/180 (103%)	180 (97%)	6 (3%)	0	100	100
1	B	182/180 (101%)	171 (94%)	8 (4%)	3 (2%)	7	1
All	All	368/360 (102%)	351 (95%)	14 (4%)	3 (1%)	14	4

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	757	SER
1	B	758	GLN
1	B	755	PRO

### 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	178/171 (104%)	178 (100%)	0	100	100
1	B	165/171 (96%)	161 (98%)	4 (2%)	43	20
All	All	343/342 (100%)	339 (99%)	4 (1%)	78	45

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	850[A]	GLU
1	B	850[B]	GLU
1	B	862[A]	ASP
1	B	862[B]	ASP

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

Mol	Chain	Res	Type
1	A	758	GLN
1	A	867	ASN
1	A	870	ASN
1	B	701	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 oligosaccharides in this entry.

### 5.6 Ligand geometry [i](#)

Of 8 ligands modelled in this entry, 5 are monoatomic - leaving 3 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
4	EDO	A	905	-	3,3,3	0.63	0	2,2,2	0.06	0
3	MES	A	904	-	12,12,12	1.76	2 (16%)	15,16,16	2.11	5 (33%)
4	EDO	B	903	-	3,3,3	0.33	0	2,2,2	0.21	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
4	EDO	A	905	-	-	1/1/1/1	-
3	MES	A	904	-	-	1/6/14/14	0/1/1/1
4	EDO	B	903	-	-	1/1/1/1	-

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	904	MES	C8-S	-5.31	1.70	1.77
3	A	904	MES	O1S-S	2.10	1.51	1.45

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	904	MES	C6-C5-N4	4.51	116.98	110.12
3	A	904	MES	C6-O1-C2	3.36	120.74	109.88
3	A	904	MES	O1S-S-C8	3.19	111.55	106.73
3	A	904	MES	C7-N4-C3	2.76	118.59	111.24
3	A	904	MES	O1-C6-C5	2.74	117.67	111.77

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	904	MES	C8-C7-N4-C3
4	A	905	EDO	O1-C1-C2-O2
4	B	903	EDO	O1-C1-C2-O2

There are no ring outliers.

3 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	905	EDO	1	0
3	A	904	MES	2	0
4	B	903	EDO	1	0

## 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	178/180 (98%)	0.38	9 (5%) 33 37	12, 24, 57, 93	10 (5%)
1	B	177/180 (98%)	0.71	22 (12%) 8 9	13, 25, 81, 136	7 (3%)
All	All	355/360 (98%)	0.54	31 (8%) 16 17	12, 24, 73, 136	17 (4%)

All (31) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	711	TRP	4.9
1	B	757	SER	4.8
1	A	735	ALA	4.8
1	B	758	GLN	4.7
1	B	800	TYR	4.4
1	B	806	SER	4.1
1	B	759	SER	4.0
1	B	756	GLU	3.5
1	B	805	GLY	3.5
1	A	736	ASN	3.4
1	B	807	GLN	3.4
1	A	738	ASN	3.3
1	B	755	PRO	3.2
1	B	804	GLU	3.1
1	B	760	GLY	3.0
1	B	754	CYS	3.0
1	B	801	TYR	2.8
1	B	803	ARG	2.7
1	B	737	LYS	2.7
1	B	876	GLU	2.6
1	A	878	SER	2.6
1	B	736	ASN	2.5
1	B	701	ASN	2.4
1	B	735	ALA	2.4

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	739	PRO	2.4
1	B	874	ILE	2.3
1	A	750	ILE	2.3
1	B	738	ASN	2.2
1	A	737	LYS	2.2
1	A	733	VAL	2.1
1	A	753	ARG	2.1

## 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 oligosaccharides 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(Å <sup>2</sup> )	Q<0.9
4	EDO	A	905	4/4	0.71	0.19	57,67,71,72	0
4	EDO	B	903	4/4	0.93	0.09	36,41,41,43	0
3	MES	A	904	12/12	0.94	0.13	21,28,34,34	12
2	ZN	A	903	1/1	0.95	0.10	38,38,38,38	1
2	ZN	A	902	1/1	0.95	0.06	31,31,31,31	0
2	ZN	A	901	1/1	0.96	0.05	24,24,24,24	0
2	ZN	B	901	1/1	0.98	0.03	19,19,19,19	0
2	ZN	B	902	1/1	0.99	0.03	23,23,23,23	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.