



# Full wwPDB X-ray Structure Validation Report ⓘ

May 24, 2020 – 11:00 pm BST

PDB ID : 4ZOA  
Title : Crystal Structure of beta-glucosidase from *Listeria innocua* in complex with isofagomine  
Authors : Nakajima, M.; Yoshida, R.; Miyanaga, A.; Abe, K.; Takahashi, Y.; Sugimoto, N.; Toyozumi, H.; Nakai, H.; Kitaoka, M.; Taguchi, H.  
Deposited on : 2015-05-06  
Resolution : 2.17 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 : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.11  
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) : 2.11



## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 11711 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.

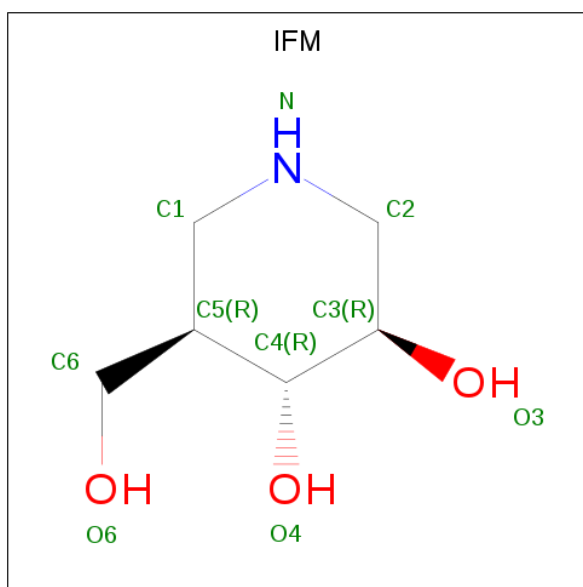
- Molecule 1 is a protein called Lin1840 protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	716	5536	3507	916	1089	24	0	0	0
1	B	715	5527	3501	914	1088	24	0	0	0

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	2	GLU	LYS	engineered mutation	UNP Q92AS9
A	724	LEU	-	expression tag	UNP Q92AS9
A	725	GLU	-	expression tag	UNP Q92AS9
A	726	HIS	-	expression tag	UNP Q92AS9
A	727	HIS	-	expression tag	UNP Q92AS9
A	728	HIS	-	expression tag	UNP Q92AS9
A	729	HIS	-	expression tag	UNP Q92AS9
A	730	HIS	-	expression tag	UNP Q92AS9
A	731	HIS	-	expression tag	UNP Q92AS9
B	2	GLU	LYS	engineered mutation	UNP Q92AS9
B	724	LEU	-	expression tag	UNP Q92AS9
B	725	GLU	-	expression tag	UNP Q92AS9
B	726	HIS	-	expression tag	UNP Q92AS9
B	727	HIS	-	expression tag	UNP Q92AS9
B	728	HIS	-	expression tag	UNP Q92AS9
B	729	HIS	-	expression tag	UNP Q92AS9
B	730	HIS	-	expression tag	UNP Q92AS9
B	731	HIS	-	expression tag	UNP Q92AS9

- Molecule 2 is 5-HYDROXYMETHYL-3,4-DIHYDROXYPIPERIDINE (three-letter code: IFM) (formula: C<sub>6</sub>H<sub>13</sub>NO<sub>3</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	
			Total	C	N			O
2	A	1	10	6	1	3	0	0
2	B	1	10	6	1	3	0	0

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

- Molecule 4 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: C<sub>4</sub>H<sub>10</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	C O	0	0
			7	4 3		

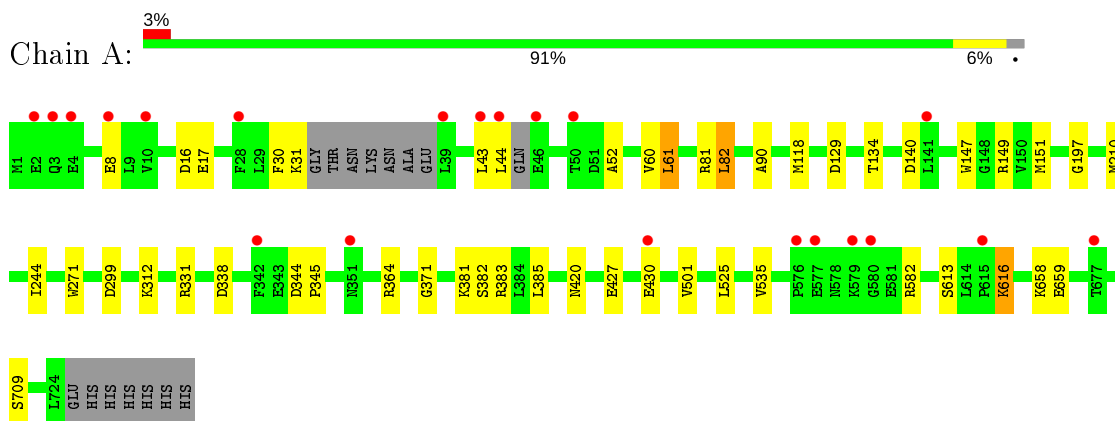
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	315	Total	O	0	0
			315	315		
5	B	304	Total	O	0	0
			304	304		

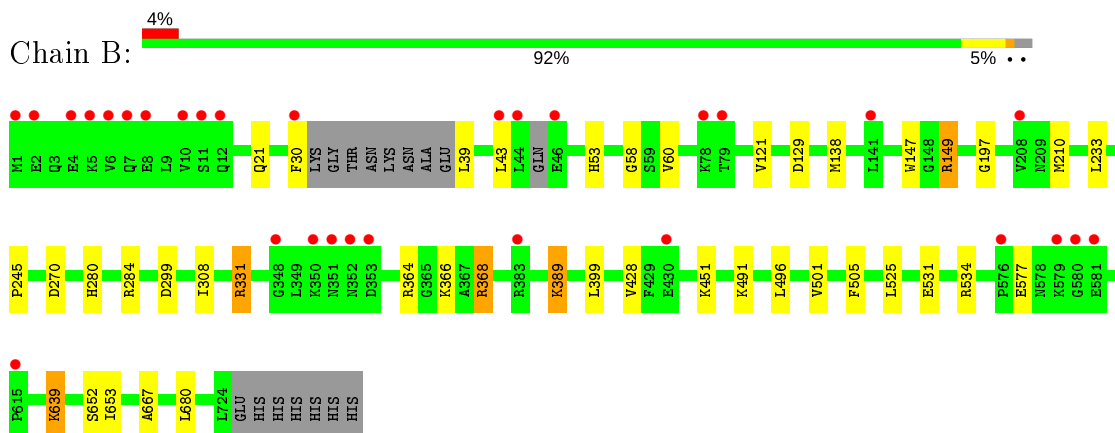
### 3 Residue-property plots [i](#)

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: Lin1840 protein



- Molecule 1: Lin1840 protein



## 4 Data and refinement statistics

Property	Value	Source
Space group	I 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	89.65Å 94.85Å 215.87Å 90.00° 96.20° 90.00°	Depositor
Resolution (Å)	40.71 – 2.17 40.71 – 2.17	Depositor EDS
% Data completeness (in resolution range)	97.6 (40.71-2.17) 97.6 (40.71-2.17)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.28 (at 2.18Å)	Xtrriage
Refinement program	REFMAC 5.7.0029	Depositor
R, $R_{free}$	0.191 , 0.258 0.200 , 0.264	Depositor DCC
$R_{free}$ test set	4638 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	17.5	Xtrriage
Anisotropy	0.177	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 45.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	11711	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	20.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 13.83% 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: IFM, MG, PEG

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	0.89	1/5636 (0.0%)	0.90	6/7636 (0.1%)
1	B	0.85	0/5627	0.91	9/7625 (0.1%)
All	All	0.87	1/11263 (0.0%)	0.90	15/15261 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	659	GLU	CD-OE2	-6.14	1.18	1.25

All (15) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	284	ARG	NE-CZ-NH1	9.98	125.29	120.30
1	B	364	ARG	NE-CZ-NH1	7.19	123.89	120.30
1	B	364	ARG	NE-CZ-NH2	-7.07	116.77	120.30
1	B	368	ARG	NE-CZ-NH2	6.51	123.55	120.30
1	A	118	MET	CG-SD-CE	5.74	109.38	100.20
1	A	364	ARG	NE-CZ-NH2	-5.71	117.44	120.30
1	A	338	ASP	CB-CG-OD1	5.70	123.43	118.30
1	B	501	VAL	CB-CA-C	-5.52	100.92	111.40
1	B	149	ARG	NE-CZ-NH2	-5.36	117.62	120.30
1	B	639	LYS	CD-CE-NZ	-5.32	99.47	111.70
1	B	331	ARG	NE-CZ-NH1	5.31	122.96	120.30
1	A	140	ASP	CB-CG-OD1	5.27	123.04	118.30
1	A	525	LEU	CA-CB-CG	5.18	127.22	115.30
1	B	368	ARG	NE-CZ-NH1	-5.17	117.71	120.30
1	A	364	ARG	NE-CZ-NH1	5.16	122.88	120.30

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	5536	0	5489	25	0
1	B	5527	0	5476	23	0
2	A	10	0	13	0	0
2	B	10	0	13	1	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	B	7	0	10	0	0
5	A	315	0	0	4	0
5	B	304	0	0	1	0
All	All	11711	0	11001	44	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 2.

All (44) 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:299:ASP:OD1	1:A:331:ARG:NH1	2.09	0.86
1:A:582:ARG:NH2	5:A:902:HOH:O	2.18	0.77
1:A:244:ILE:HD11	1:B:652:SER:O	1.94	0.68
1:B:121:VAL:HG21	1:B:366:LYS:HD2	1.75	0.67
1:A:244:ILE:HD11	1:B:653:ILE:HD13	1.78	0.64
1:A:52:ALA:HB3	5:A:1131:HOH:O	2.00	0.62
1:B:39:LEU:HD11	1:B:43:LEU:HD23	1.82	0.62
1:A:299:ASP:CG	1:A:331:ARG:HH11	2.04	0.60
1:A:427:GLU:OE2	5:A:901:HOH:O	2.16	0.59
1:A:244:ILE:CD1	1:B:652:SER:O	2.54	0.55
1:A:658:LYS:HE2	1:A:709:SER:O	2.06	0.55
1:B:368:ARG:NH1	1:B:534:ARG:HD3	2.22	0.54
1:A:147:TRP:CZ2	1:A:149:ARG:HD3	2.43	0.54
1:B:639:LYS:HG2	1:B:667:ALA:HA	1.92	0.52
1:B:30:PHE:CE2	1:B:60:VAL:HG22	2.46	0.51
1:B:299:ASP:OD1	1:B:331:ARG:NH1	2.44	0.50
1:B:451:LYS:HD2	1:B:496:LEU:HD22	1.93	0.50
1:A:381:LYS:O	1:A:382:SER:HB2	2.12	0.50

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:331:ARG:NH2	5:B:901:HOH:O	2.31	0.49
1:A:61:LEU:HD23	1:A:61:LEU:O	2.12	0.49
1:A:81:ARG:NH2	1:A:82:LEU:HD21	2.28	0.48
1:B:389:LYS:HE3	1:B:428:VAL:O	2.14	0.48
1:B:147:TRP:CZ3	1:B:197:GLY:HA2	2.48	0.48
1:B:270:ASP:OD1	2:B:802:IFM:N	2.47	0.47
1:B:531:GLU:OE1	1:B:534:ARG:HD2	2.14	0.47
1:B:505:PHE:CZ	1:B:525:LEU:HD13	2.50	0.46
1:A:616:LYS:HD3	1:A:616:LYS:O	2.16	0.46
1:B:53:HIS:CD2	1:B:308:ILE:HD12	2.50	0.46
1:A:16:ASP:OD1	1:A:312:LYS:NZ	2.47	0.45
1:A:271:TRP:CD1	1:A:271:TRP:C	2.90	0.45
1:A:385:LEU:HD21	1:A:501:VAL:HG21	1.99	0.44
1:B:577:GLU:CD	1:B:577:GLU:H	2.20	0.43
1:A:420:ASN:HB3	5:A:1176:HOH:O	2.19	0.43
1:A:244:ILE:CD1	1:B:653:ILE:HD13	2.47	0.43
1:A:30:PHE:CE1	1:A:60:VAL:HG22	2.54	0.43
1:A:90:ALA:O	1:A:134:THR:HA	2.20	0.42
1:A:147:TRP:CZ3	1:A:197:GLY:HA2	2.54	0.42
1:B:233:LEU:HD12	1:B:233:LEU:N	2.34	0.42
1:B:245:PRO:HG3	1:B:280:HIS:CG	2.55	0.42
1:A:17:GLU:OE2	1:A:81:ARG:NH1	2.53	0.41
1:A:344:ASP:HA	1:A:345:PRO:HD2	1.88	0.41
1:B:21:GLN:O	1:B:58:GLY:HA3	2.20	0.41
1:B:147:TRP:CZ2	1:B:149:ARG:HD3	2.56	0.41
1:A:371:GLY:HA3	1:A:535:VAL:HG21	2.02	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	710/731 (97%)	691 (97%)	19 (3%)	0	100	100
1	B	709/731 (97%)	689 (97%)	20 (3%)	0	100	100
All	All	1419/1462 (97%)	1380 (97%)	39 (3%)	0	100	100

There are no Ramachandran outliers to report.

### 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	599/612 (98%)	586 (98%)	13 (2%)	52	62
1	B	598/612 (98%)	591 (99%)	7 (1%)	71	81
All	All	1197/1224 (98%)	1177 (98%)	20 (2%)	60	72

All (20) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	8	GLU
1	A	31	LYS
1	A	43	LEU
1	A	44	LEU
1	A	61	LEU
1	A	82	LEU
1	A	129	ASP
1	A	151	MET
1	A	210	MET
1	A	383	ARG
1	A	430	GLU
1	A	613	SER
1	A	616	LYS
1	B	129	ASP
1	B	138	MET
1	B	210	MET
1	B	389	LYS
1	B	399	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	491	LYS
1	B	680	LEU

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	455	GLN

### 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](#)

Of 5 ligands modelled in this entry, 2 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	$\# Z  > 2$
2	IFM	A	801	-	9,10,10	1.23	1 (11%)	9,13,13	2.27	3 (33%)
4	PEG	B	801	-	6,6,6	0.65	0	5,5,5	0.68	0
2	IFM	B	802	-	9,10,10	1.81	1 (11%)	9,13,13	2.36	1 (11%)

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	IFM	A	801	-	-	0/2/16/16	0/1/1/1
4	PEG	B	801	-	-	4/4/4/4	-
2	IFM	B	802	-	-	0/2/16/16	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	802	IFM	C2-C3	4.77	1.57	1.52
2	A	801	IFM	C2-C3	3.02	1.55	1.52

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	802	IFM	C1-N-C2	6.58	118.92	111.70
2	A	801	IFM	C1-N-C2	4.57	116.72	111.70
2	A	801	IFM	C2-C3-C4	3.70	114.67	110.33
2	A	801	IFM	O3-C3-C2	2.35	114.21	109.61

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	B	801	PEG	O1-C1-C2-O2
4	B	801	PEG	O2-C3-C4-O4
4	B	801	PEG	C1-C2-O2-C3
4	B	801	PEG	C4-C3-O2-C2

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	802	IFM	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	716/731 (97%)	0.15	21 (2%) 51 52	7, 17, 37, 60	0
1	B	715/731 (97%)	0.23	30 (4%) 36 37	8, 18, 40, 68	0
All	All	1431/1462 (97%)	0.19	51 (3%) 42 43	7, 18, 39, 68	0

All (51) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	351	ASN	5.4
1	B	580	GLY	4.6
1	A	44	LEU	4.5
1	A	576	PRO	4.3
1	B	44	LEU	4.2
1	B	579	LYS	4.1
1	A	579	LYS	4.0
1	B	348	GLY	3.9
1	A	43	LEU	3.8
1	B	11	SER	3.4
1	B	46	GLU	3.4
1	A	46	GLU	3.3
1	A	2	GLU	3.2
1	B	576	PRO	3.2
1	B	350	LYS	3.2
1	B	2	GLU	3.1
1	A	351	ASN	3.0
1	A	10	VAL	3.0
1	B	7	GLN	2.9
1	A	4	GLU	2.9
1	B	5	LYS	2.8
1	A	580	GLY	2.7
1	B	430	GLU	2.7
1	A	342	PHE	2.6

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	B	353	ASP	2.5
1	A	39	LEU	2.5
1	B	8	GLU	2.5
1	B	12	GLN	2.5
1	A	430	GLU	2.5
1	A	28	PHE	2.4
1	A	50	THR	2.4
1	A	8	GLU	2.3
1	B	141	LEU	2.3
1	B	78	LYS	2.3
1	B	43	LEU	2.3
1	B	352	ASN	2.3
1	B	6	VAL	2.3
1	B	1	MET	2.2
1	B	30	PHE	2.2
1	B	581	GLU	2.2
1	A	3	GLN	2.1
1	B	208	VAL	2.1
1	B	4	GLU	2.1
1	B	79	THR	2.1
1	B	615	PRO	2.1
1	A	141	LEU	2.1
1	A	577	GLU	2.1
1	A	615	PRO	2.0
1	B	383	ARG	2.0
1	B	10	VAL	2.0
1	A	677	THR	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(Å <sup>2</sup> )	Q<0.9
4	PEG	B	801	7/7	0.89	0.15	30,34,35,37	0
3	MG	A	802	1/1	0.97	0.04	19,19,19,19	0
2	IFM	A	801	10/10	0.97	0.11	7,9,10,10	0
2	IFM	B	802	10/10	0.97	0.11	8,11,11,12	0
3	MG	B	803	1/1	0.98	0.03	21,21,21,21	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.