wwPDB X-ray Structure Validation Summary Report (i)

Aug 21, 2020 – 02:29 AM BST

PDB ID	:	5AZD
Title	:	Crystal structure of thermophilic rhodopsin.
Authors	:	Mizutani, K.; Hashimoto, N.; Tsukamoto, T.; Yamashita, K.; Yamamoto, M.; Sudo, Y.; Murata, T.
Deposited on	:	2015-09-30
Resolution	:	2.80 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

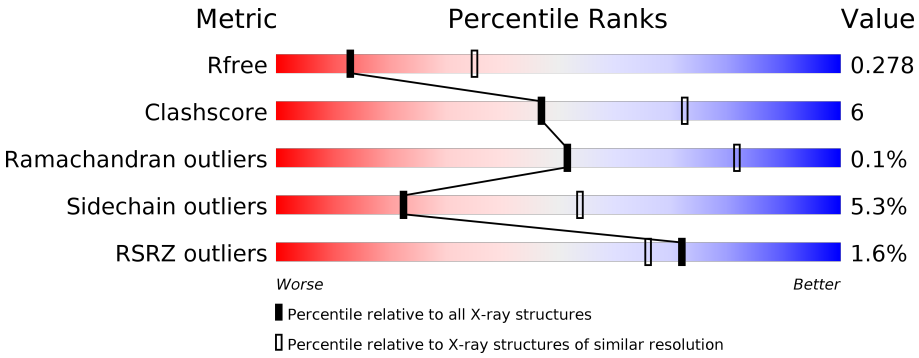
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.13.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.13.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.80 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	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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 >=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 <=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	268	74% 19% • 6%
1	B	268	74% 18% • 7%
1	C	268	79% 13% • 6%
1	D	268	79% 13% • 7%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 7900 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 Bacteriorhodopsin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	251	Total	C	N	O	S	0	0	0
			1986	1348	306	327	5			
1	B	250	Total	C	N	O	S	0	0	0
			1975	1342	302	326	5			
1	C	251	Total	C	N	O	S	0	0	0
			1972	1343	300	325	4			
1	D	250	Total	C	N	O	S	0	0	0
			1965	1334	302	325	4			

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	261	LEU	-	expression tag	UNP H9ZSC3
A	262	GLU	-	expression tag	UNP H9ZSC3
A	263	HIS	-	expression tag	UNP H9ZSC3
A	264	HIS	-	expression tag	UNP H9ZSC3
A	265	HIS	-	expression tag	UNP H9ZSC3
A	266	HIS	-	expression tag	UNP H9ZSC3
A	267	HIS	-	expression tag	UNP H9ZSC3
A	268	HIS	-	expression tag	UNP H9ZSC3
B	261	LEU	-	expression tag	UNP H9ZSC3
B	262	GLU	-	expression tag	UNP H9ZSC3
B	263	HIS	-	expression tag	UNP H9ZSC3
B	264	HIS	-	expression tag	UNP H9ZSC3
B	265	HIS	-	expression tag	UNP H9ZSC3
B	266	HIS	-	expression tag	UNP H9ZSC3
B	267	HIS	-	expression tag	UNP H9ZSC3
B	268	HIS	-	expression tag	UNP H9ZSC3
C	261	LEU	-	expression tag	UNP H9ZSC3
C	262	GLU	-	expression tag	UNP H9ZSC3
C	263	HIS	-	expression tag	UNP H9ZSC3
C	264	HIS	-	expression tag	UNP H9ZSC3
C	265	HIS	-	expression tag	UNP H9ZSC3

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	266	HIS	-	expression tag	UNP H9ZSC3
C	267	HIS	-	expression tag	UNP H9ZSC3
C	268	HIS	-	expression tag	UNP H9ZSC3
D	261	LEU	-	expression tag	UNP H9ZSC3
D	262	GLU	-	expression tag	UNP H9ZSC3
D	263	HIS	-	expression tag	UNP H9ZSC3
D	264	HIS	-	expression tag	UNP H9ZSC3
D	265	HIS	-	expression tag	UNP H9ZSC3
D	266	HIS	-	expression tag	UNP H9ZSC3
D	267	HIS	-	expression tag	UNP H9ZSC3
D	268	HIS	-	expression tag	UNP H9ZSC3

- Molecule 2 is water.

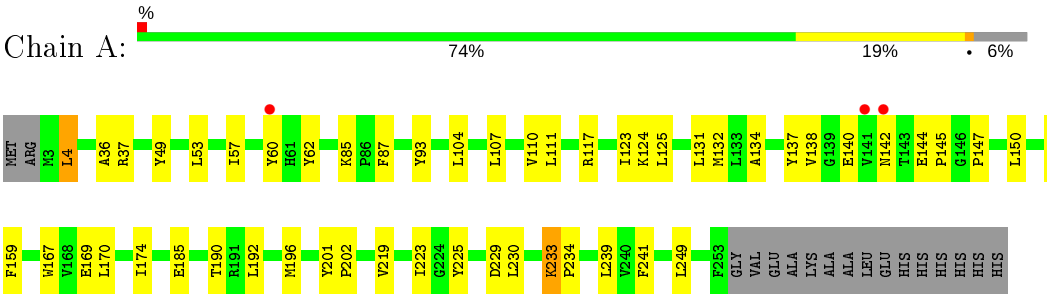
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total O 1 1	0	0
2	B	1	Total O 1 1	0	0

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.

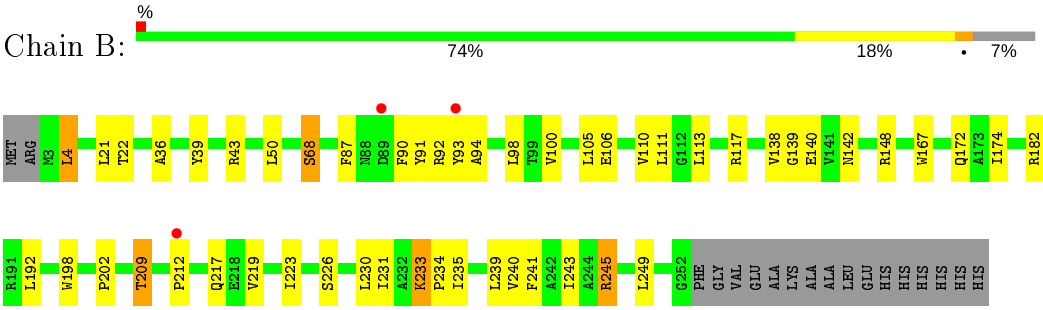
- Molecule 1: Bacteriorhodopsin

Chain A:



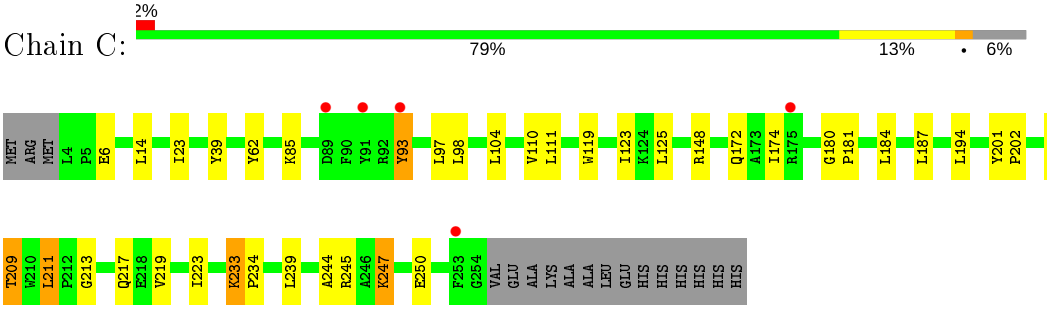
- Molecule 1: Bacteriorhodopsin

Chain B:



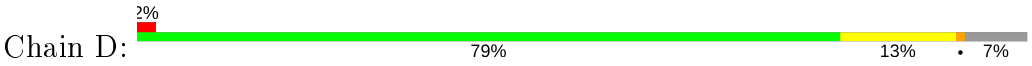
- Molecule 1: Bacteriorhodopsin

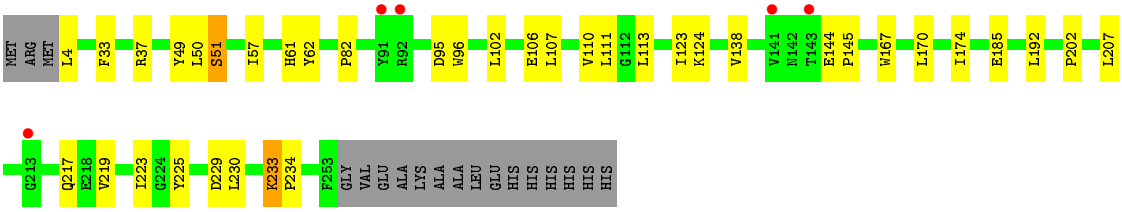
Chain C:



- Molecule 1: Bacteriorhodopsin

Chain D:




MET ARG MET L4 F33 R37 Y49 L50 S51 I57 H61 Y62 P82 Y91 R92 D95 W96 L102 E106 L107 V110 L111 G112 L113 I123 K124 V138 V141 N142 T143 E144 P145 W167 L170 I174 E185 L192 P202 L207
G213 Q217 E218 V219 I223 G224 Y225 D229 L230 K233 P234 F253 GLY VAL GLU ALA LYS ALA ALA LEU GLU HIS HIS HIS HIS HIS HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	63.19Å 67.69Å 73.05Å 116.48° 114.14° 90.82°	Depositor
Resolution (Å)	46.52 – 2.80 46.52 – 2.80	Depositor EDS
% Data completeness (in resolution range)	93.3 (46.52-2.80) 93.2 (46.52-2.80)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.00 (at 2.81Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.228 , 0.278 0.230 , 0.278	Depositor DCC
R_{free} test set	1129 reflections (5.10%)	wwPDB-VP
Wilson B-factor (Å^2)	39.2	Xtriage
Anisotropy	0.001	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 60.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.016 for h,-k,-h-l 0.000 for -h,k,-k-l 0.000 for -h,-k,h+k+l	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	7900	wwPDB-VP
Average B, all atoms (Å^2)	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 47.93 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 9.2518e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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: LYR

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	0.23	0/2014	0.37	0/2755
1	B	0.24	0/2003	0.39	0/2741
1	C	0.24	0/2001	0.39	0/2740
1	D	0.22	0/1992	0.36	0/2727
All	All	0.23	0/8010	0.38	0/10963

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

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	90	PHE	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	1986	0	2037	27	0
1	B	1975	0	2024	28	0
1	C	1972	0	2012	25	0
1	D	1965	0	2010	21	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
All	All	7900	0	8083	99	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 6.

The worst 5 of 99 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:124:LYS:HE2	1:A:169:GLU:OE2	1.90	0.70
1:A:190:THR:HG22	1:A:239:LEU:HB3	1.77	0.66
1:C:111:LEU:HD13	1:C:174:ILE:HD11	1.80	0.64
1:C:180:GLY:H	1:C:184:LEU:HD13	1.63	0.64
1:B:212:PRO:O	1:B:217:GLN:NE2	2.32	0.62

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	248/268 (92%)	245 (99%)	3 (1%)	0	100 100
1	B	247/268 (92%)	244 (99%)	2 (1%)	1 (0%)	34 66
1	C	248/268 (92%)	241 (97%)	7 (3%)	0	100 100
1	D	247/268 (92%)	245 (99%)	2 (1%)	0	100 100
All	All	990/1072 (92%)	975 (98%)	14 (1%)	1 (0%)	51 81

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	92	ARG

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	196/214 (92%)	184 (94%)	12 (6%)	18	48
1	B	195/214 (91%)	183 (94%)	12 (6%)	18	47
1	C	193/214 (90%)	184 (95%)	9 (5%)	26	59
1	D	193/214 (90%)	185 (96%)	8 (4%)	30	64
All	All	777/856 (91%)	736 (95%)	41 (5%)	22	54

5 of 41 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	167	TRP
1	B	245	ARG
1	D	192	LEU
1	B	172	GLN
1	B	182	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	142	ASN
1	D	76	GLN
1	B	217	GLN
1	A	142	ASN
1	C	217	GLN

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

4 non-standard protein/DNA/RNA residues 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	$\#\|Z\|>2$
1	LYR	D	233	1	27,29,30	0.69	0	30,37,39	1.10	2 (6%)
1	LYR	C	233	1	27,29,30	0.65	0	30,37,39	0.92	1 (3%)
1	LYR	B	233	1	27,29,30	0.63	0	30,37,39	1.04	1 (3%)
1	LYR	A	233	1	27,29,30	0.73	0	30,37,39	0.76	1 (3%)

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
1	LYR	D	233	1	-	9/22/40/42	0/1/1/1
1	LYR	C	233	1	-	10/22/40/42	0/1/1/1
1	LYR	B	233	1	-	8/22/40/42	0/1/1/1
1	LYR	A	233	1	-	11/22/40/42	0/1/1/1

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	233	LYR	C4-C3-C5	-3.20	113.04	118.08
1	D	233	LYR	C4-C3-C5	-2.97	113.41	118.08
1	B	233	LYR	C4-C3-C5	-2.62	113.95	118.08
1	D	233	LYR	C16-C17-C11	2.58	114.46	110.48

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	233	LYR	C4-C3-C5	-2.42	114.27	118.08

There are no chirality outliers.

5 of 38 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	C	233	LYR	C9-C10-C11-C12
1	B	233	LYR	C8-C80-C9-C10
1	A	233	LYR	C4-C3-C5-C6
1	A	233	LYR	C5-C6-C7-C80
1	A	233	LYR	C7-C80-C9-C10

There are no ring outliers.

4 monomers are involved in 11 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	D	233	LYR	3	0
1	C	233	LYR	3	0
1	B	233	LYR	3	0
1	A	233	LYR	2	0

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

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^{th} 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^2)	Q<0.9
1	A	250/268 (93%)	-0.42	3 (1%)	79	73	27, 42, 72, 97	0
1	B	249/268 (92%)	-0.41	3 (1%)	79	73	26, 42, 77, 105	0
1	C	250/268 (93%)	-0.34	5 (2%)	65	56	26, 45, 81, 124	0
1	D	249/268 (92%)	-0.42	5 (2%)	65	56	24, 40, 73, 108	0
All	All	998/1072 (93%)	-0.39	16 (1%)	72	66	24, 42, 76, 124	0

The worst 5 of 16 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	93	TYR	3.8
1	D	91	TYR	3.5
1	D	213	GLY	2.9
1	C	253	PHE	2.9
1	D	92	ARG	2.9

6.2 Non-standard residues in protein, DNA, RNA chains

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^{th} 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^2)	Q<0.9
1	LYR	C	233	29/30	0.89	0.18	32,37,43,45	0
1	LYR	B	233	29/30	0.89	0.15	24,30,36,36	0
1	LYR	A	233	29/30	0.92	0.17	27,30,34,34	0
1	LYR	D	233	29/30	0.93	0.17	27,34,36,36	0

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.