



# Full wwPDB X-ray Structure Validation Report ⓘ

Nov 19, 2017 – 02:15 AM EST

PDB ID : 5II5  
Title : Crystal structure of red abalone VERL repeat 1 at 1.8 Å resolution  
Authors : Sadat Al-Hosseini, H.; Raj, I.; Nishimura, K.; Jovine, L.  
Deposited on : unknown  
Resolution : 1.80 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

<http://wwpdb.org/validation/2016/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 : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20030345  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20030345

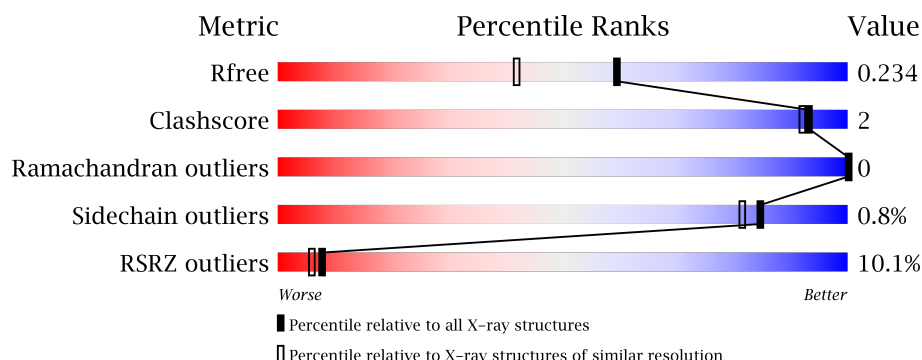
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.80 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))
$R_{free}$	100719	4827 (1.80-1.80)
Clashscore	112137	5742 (1.80-1.80)
Ramachandran outliers	110173	5676 (1.80-1.80)
Sidechain outliers	110143	5675 (1.80-1.80)
RSRZ outliers	101464	4906 (1.80-1.80)

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  $\geq 3$ , 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	496	<div> <div>10%</div> <div>91%</div> <div>5%</div> <div>.</div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	MAL	A	4201	-	-	-	X
3	PGE	A	4202	-	-	-	X

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	PGE	A	4204	-	-	-	X
3	PGE	A	4205	-	-	-	X

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 7668 atoms, of which 3683 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 Maltose-binding periplasmic protein, Vitelline envelope sperm lysin receptor.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	476	Total	C	H	N	O	S	0	3	0
			7310	2378	3606	606	705	15			

There are 31 discrepancies between the modelled and reference sequences:

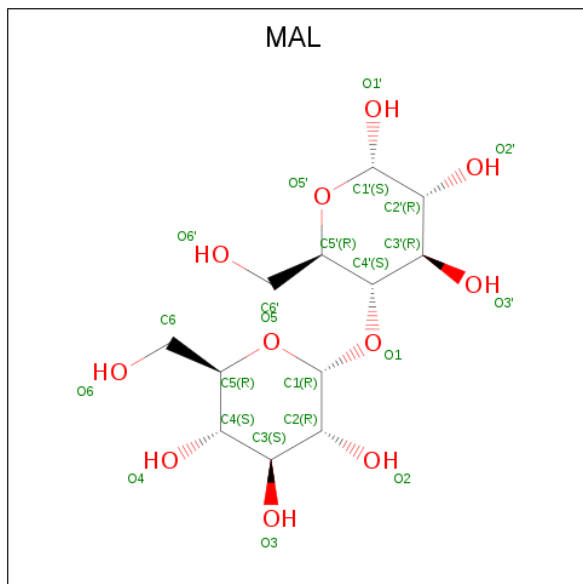
Chain	Residue	Modelled	Actual	Comment	Reference
A	3664	GLU	-	expression tag	UNP P0AEX9
A	3665	THR	-	expression tag	UNP P0AEX9
A	3666	GLY	-	expression tag	UNP P0AEX9
A	3667	THR	-	expression tag	UNP P0AEX9
A	3749	ALA	ASP	conflict	UNP P0AEX9
A	3750	ALA	LYS	conflict	UNP P0AEX9
A	3839	ALA	GLU	conflict	UNP P0AEX9
A	3840	ALA	ASN	conflict	UNP P0AEX9
A	3882	HIS	ALA	conflict	UNP P0AEX9
A	3886	HIS	LYS	conflict	UNP P0AEX9
A	3906	ALA	LYS	conflict	UNP P0AEX9
A	3979	VAL	ALA	conflict	UNP P0AEX9
A	3984	VAL	ILE	conflict	UNP P0AEX9
A	4026	ALA	GLU	conflict	UNP P0AEX9
A	4029	ALA	LYS	conflict	UNP P0AEX9
A	4030	ALA	ASP	conflict	UNP P0AEX9
A	4034	ASN	ARG	conflict	UNP P0AEX9
A	4035	ALA	-	linker	UNP P0AEX9
A	4036	ALA	-	linker	UNP P0AEX9
A	4037	ALA	-	linker	UNP P0AEX9
A	4115	GLN	ASN	conflict	UNP Q8WR62
A	4122	THR	ASN	conflict	UNP Q8WR62
A	4142	TYR	ASN	conflict	UNP Q8WR62
A	4152	LEU	-	expression tag	UNP Q8WR62
A	4153	GLU	-	expression tag	UNP Q8WR62
A	4154	HIS	-	expression tag	UNP Q8WR62

*Continued on next page...*

Continued from previous page...

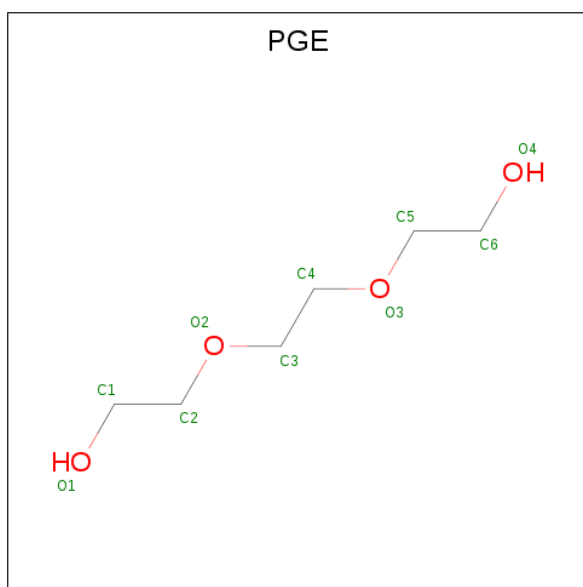
Chain	Residue	Modelled	Actual	Comment	Reference
A	4155	HIS	-	expression tag	UNP Q8WR62
A	4156	HIS	-	expression tag	UNP Q8WR62
A	4157	HIS	-	expression tag	UNP Q8WR62
A	4158	HIS	-	expression tag	UNP Q8WR62
A	4159	HIS	-	expression tag	UNP Q8WR62

- Molecule 2 is MALTOSE (three-letter code: MAL) (formula:  $C_{12}H_{22}O_{11}$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	H	O	0	0
			44	12	21	11		

- Molecule 3 is TRIETHYLENE GLYCOL (three-letter code: PGE) (formula:  $C_6H_{14}O_4$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	H	O	0	0
			24	6	14	4		
3	A	1	Total	C	H	O	0	0
			24	6	14	4		
3	A	1	Total	C	H	O	0	0
			24	6	14	4		
3	A	1	Total	C	H	O	0	0
			24	6	14	4		

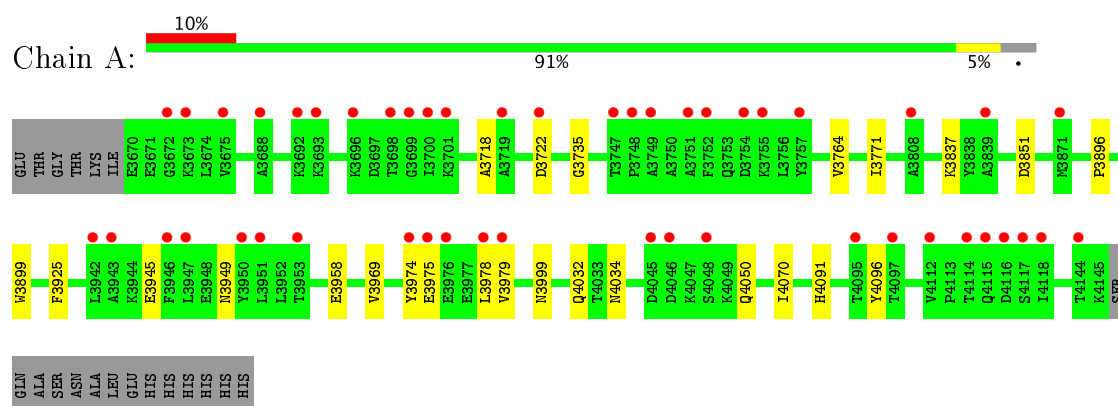
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	217	Total	O	0	1
			218	218		

### 3 Residue-property plots

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: Maltose-binding periplasmic protein,Vitelline envelope sperm lysin receptor



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 65 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	107.09Å 107.09Å 196.35Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	37.84 – 1.80 37.84 – 1.80	Depositor EDS
% Data completeness (in resolution range)	99.2 (37.84-1.80) 94.0 (37.84-1.80)	Depositor EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.81 (at 1.81Å)	Xtriage
Refinement program	PHENIX dev-1894_1692	Depositor
R, $R_{free}$	0.198 , 0.233 0.196 , 0.234	Depositor DCC
$R_{free}$ test set	2978 reflections (5.10%)	DCC
Wilson B-factor (Å <sup>2</sup> )	29.7	Xtriage
Anisotropy	0.441	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.39 , 56.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	7668	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	64.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.81% 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: PGE, MAL

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	0.55	0/3797	0.63	0/5168

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 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	3704	3606	3640	13	0
2	A	23	21	22	0	0
3	A	40	56	56	0	0
4	A	218	0	0	1	0
All	All	3985	3683	3718	13	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 (13) 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:3945:GLU:OE1	1:A:3949:ASN:ND2	2.33	0.61
1:A:3851:ASP:HB2	1:A:4032:GLN:HB2	1.86	0.57
1:A:3975:GLU:O	1:A:3979:VAL:HG23	2.07	0.54
1:A:4070:ILE:HD12	1:A:4070:ILE:N	2.30	0.45
1:A:3896:PRO:HA	1:A:3899:TRP:CE2	2.51	0.45
1:A:3958:GLU:HG3	1:A:3974:TYR:CE2	2.52	0.45
1:A:3969:VAL:HG21	1:A:3974:TYR:HB3	2.00	0.43
1:A:3718:ALA:HA	1:A:3722:ASP:O	2.18	0.42
1:A:3735:GLY:HA3	1:A:3999:ASN:O	2.19	0.42
1:A:4091:HIS:HD2	1:A:4096:TYR:CE2	2.36	0.42
1:A:3764:VAL:O	1:A:3771:ILE:HG12	2.20	0.42
1:A:4050:GLN:HG3	4:A:5176:HOH:O	2.19	0.42
1:A:3974:TYR:CD1	1:A:3978:LEU:HG	2.55	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	477/496 (96%)	473 (99%)	4 (1%)	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	387/401 (96%)	384 (99%)	3 (1%)	85 81

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

Mol	Chain	Res	Type
1	A	3837	LYS
1	A	3925	PHE
1	A	4034	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	3949	ASN
1	A	4063	HIS
1	A	4091	HIS

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

5 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	MAL	A	4201	-	24,24,24	0.57	0	35,35,35	0.92	1 (2%)
3	PGE	A	4202	-	9,9,9	0.34	0	8,8,8	0.34	0
3	PGE	A	4203	-	9,9,9	0.33	0	8,8,8	0.35	0
3	PGE	A	4204	-	9,9,9	0.33	0	8,8,8	0.29	0
3	PGE	A	4205	-	9,9,9	0.31	0	8,8,8	0.36	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
2	MAL	A	4201	-	-	0/8/48/48	0/2/2/2
3	PGE	A	4202	-	-	0/7/7/7	0/0/0/0
3	PGE	A	4203	-	-	0/7/7/7	0/0/0/0
3	PGE	A	4204	-	-	0/7/7/7	0/0/0/0
3	PGE	A	4205	-	-	0/7/7/7	0/0/0/0

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	4201	MAL	O5'-C1'-C2'	2.20	113.69	110.04

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 ⓘ

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	476/496 (95%)	0.30	48 (10%) 8 6	28, 50, 104, 146	0

All (48) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	3942	LEU	8.6
1	A	3672	GLY	6.3
1	A	3699	GLY	5.7
1	A	3943	ALA	5.7
1	A	3696	LYS	5.5
1	A	3979	VAL	5.4
1	A	4116	ASP	4.4
1	A	3978	LEU	4.4
1	A	3747	THR	4.3
1	A	4118	ILE	4.2
1	A	4045	ASP	4.1
1	A	3950	TYR	3.9
1	A	3698	THR	3.8
1	A	3751	ALA	3.6
1	A	3749	ALA	3.6
1	A	3673	LYS	3.5
1	A	4112	VAL	3.4
1	A	3701	LYS	3.4
1	A	3752	PHE	3.4
1	A	3951	LEU	3.3
1	A	4117	SER	3.3
1	A	3700	ILE	3.3
1	A	3754	ASP	3.2
1	A	3946	PHE	3.2
1	A	4115	GLN	3.2
1	A	3675	VAL	3.1
1	A	3748	PRO	2.9

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	3757	TYR	2.8
1	A	3976	GLU	2.8
1	A	3974	TYR	2.7
1	A	3975	GLU	2.7
1	A	3839	ALA	2.7
1	A	3871	MET	2.7
1	A	3953	THR	2.7
1	A	4095	THR	2.7
1	A	4048	SER	2.6
1	A	3755	LYS	2.6
1	A	3808	ALA	2.5
1	A	3947	LEU	2.5
1	A	3722	ASP	2.5
1	A	4114	THR	2.4
1	A	3693	LYS	2.4
1	A	4097	THR	2.3
1	A	3692	LYS	2.3
1	A	4046	ASP	2.1
1	A	3688	ALA	2.1
1	A	3719	ALA	2.1
1	A	4144	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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
3	PGE	A	4202	10/10	0.79	0.18	7.20	67,92,124,124	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	PGE	A	4205	10/10	0.69	0.24	3.97	89,111,154,164	0
3	PGE	A	4204	10/10	0.95	0.15	3.22	64,103,129,133	0
2	MAL	A	4201	23/23	0.97	0.18	2.43	28,37,56,90	0
3	PGE	A	4203	10/10	0.92	0.12	1.84	60,96,127,138	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.