



wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 18, 2023 – 08:11 AM EST

PDB ID : 1O4Z
Title : THE THREE-DIMENSIONAL STRUCTURE OF BETA-AGARASE B FROM ZOBELLIA GALACTANIVORANS
Authors : Allouch, J.; Jam, M.; Helbert, W.; Barbeyron, T.; Kloareg, B.; Henrissat, B.; Czjzek, M.
Deposited on : 2003-07-29
Resolution : 2.30 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 ⓘ 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 : 4.02b-467
Mogul : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.36
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) : 2.36

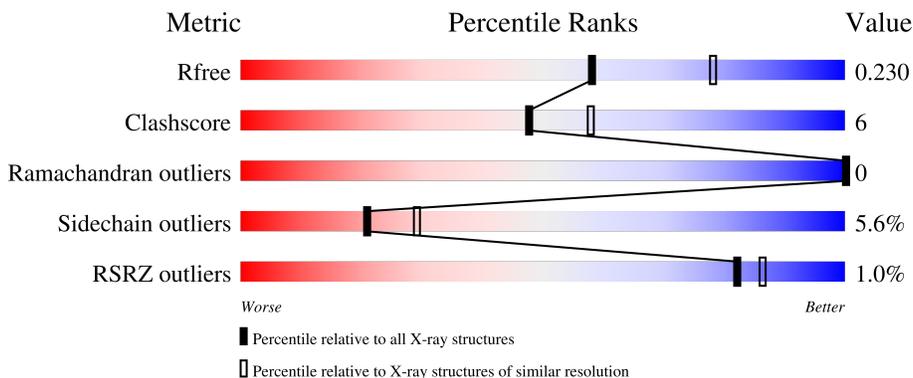
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.30 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	5042 (2.30-2.30)
Clashscore	141614	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)
RSRZ outliers	127900	4938 (2.30-2.30)

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 ≥ 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	346	 67% 16% 15%
1	B	346	 70% 14% 14%
1	C	346	 71% 14% 14%
1	D	346	 70% 14% 14%

2 Entry composition i

There are 5 unique types of molecules in this entry. The entry contains 10647 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 beta-agarase B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	295	Total 2413	C 1532	N 402	O 473	S 6	0	0	0
1	B	297	Total 2428	C 1543	N 404	O 475	S 6	0	0	0
1	C	296	Total 2421	C 1537	N 403	O 475	S 6	0	1	0
1	D	297	Total 2428	C 1543	N 404	O 475	S 6	0	0	0

There are 44 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	16	MET	-	cloning artifact	UNP Q9RGX8
A	17	ASP	-	cloning artifact	UNP Q9RGX8
A	18	ILE	-	cloning artifact	UNP Q9RGX8
A	354	LEU	-	cloning artifact	UNP Q9RGX8
A	355	GLU	-	cloning artifact	UNP Q9RGX8
A	356	HIS	-	expression tag	UNP Q9RGX8
A	357	HIS	-	expression tag	UNP Q9RGX8
A	358	HIS	-	expression tag	UNP Q9RGX8
A	359	HIS	-	expression tag	UNP Q9RGX8
A	360	HIS	-	expression tag	UNP Q9RGX8
A	361	HIS	-	expression tag	UNP Q9RGX8
B	16	MET	-	cloning artifact	UNP Q9RGX8
B	17	ASP	-	cloning artifact	UNP Q9RGX8
B	18	ILE	-	cloning artifact	UNP Q9RGX8
B	354	LEU	-	cloning artifact	UNP Q9RGX8
B	355	GLU	-	cloning artifact	UNP Q9RGX8
B	356	HIS	-	expression tag	UNP Q9RGX8
B	357	HIS	-	expression tag	UNP Q9RGX8
B	358	HIS	-	expression tag	UNP Q9RGX8
B	359	HIS	-	expression tag	UNP Q9RGX8
B	360	HIS	-	expression tag	UNP Q9RGX8

Continued on next page...

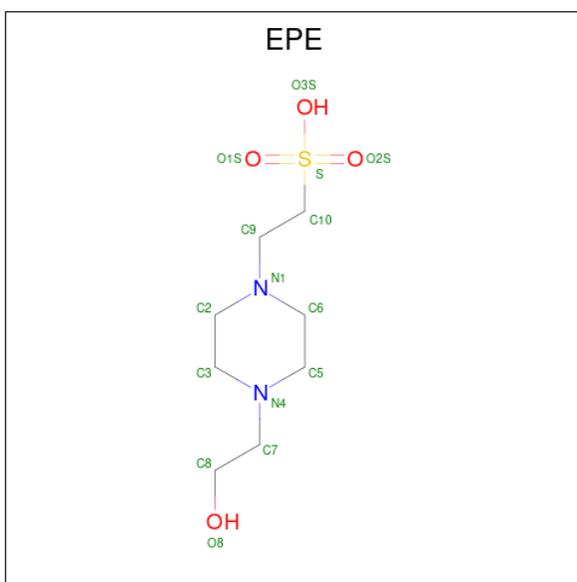
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	361	HIS	-	expression tag	UNP Q9RGX8
C	16	MET	-	cloning artifact	UNP Q9RGX8
C	17	ASP	-	cloning artifact	UNP Q9RGX8
C	18	ILE	-	cloning artifact	UNP Q9RGX8
C	354	LEU	-	cloning artifact	UNP Q9RGX8
C	355	GLU	-	cloning artifact	UNP Q9RGX8
C	356	HIS	-	expression tag	UNP Q9RGX8
C	357	HIS	-	expression tag	UNP Q9RGX8
C	358	HIS	-	expression tag	UNP Q9RGX8
C	359	HIS	-	expression tag	UNP Q9RGX8
C	360	HIS	-	expression tag	UNP Q9RGX8
C	361	HIS	-	expression tag	UNP Q9RGX8
D	16	MET	-	cloning artifact	UNP Q9RGX8
D	17	ASP	-	cloning artifact	UNP Q9RGX8
D	18	ILE	-	cloning artifact	UNP Q9RGX8
D	354	LEU	-	cloning artifact	UNP Q9RGX8
D	355	GLU	-	cloning artifact	UNP Q9RGX8
D	356	HIS	-	expression tag	UNP Q9RGX8
D	357	HIS	-	expression tag	UNP Q9RGX8
D	358	HIS	-	expression tag	UNP Q9RGX8
D	359	HIS	-	expression tag	UNP Q9RGX8
D	360	HIS	-	expression tag	UNP Q9RGX8
D	361	HIS	-	expression tag	UNP Q9RGX8

- Molecule 2 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	2	Total Na 2 2	0	0
2	B	2	Total Na 2 2	0	0
2	C	2	Total Na 2 2	0	0
2	D	1	Total Na 1 1	0	0

- Molecule 3 is 4-(2-HYDROXYETHYL)-1-PIPERAZINE ETHANESULFONIC ACID (three-letter code: EPE) (formula: C₈H₁₈N₂O₄S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	S		
3	A	1	Total 15	8	2	4	1	0	0
3	B	1	Total 15	8	2	4	1	0	0
3	C	1	Total 15	8	2	4	1	0	0
3	D	1	Total 15	8	2	4	1	0	0

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Mg		
4	B	1	Total 1	1	0	0
4	C	1	Total 1	1	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	226	Total 226	226	0	0
5	B	230	Total 230	230	0	0
5	C	213	Total 213	213	0	0

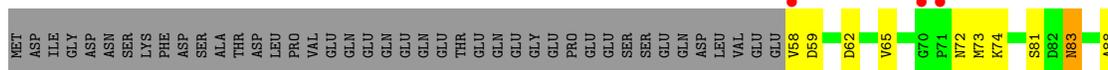
Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	D	219	Total 219	O 219	0	0



- Molecule 1: beta-agarase B



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	72.45Å 105.35Å 97.37Å 90.00° 94.13° 90.00°	Depositor
Resolution (Å)	25.00 – 2.30 24.75 – 2.30	Depositor EDS
% Data completeness (in resolution range)	99.9 (25.00-2.30) 99.9 (24.75-2.30)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.11	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.98 (at 2.31Å)	Xtrriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.166 , 0.224 0.179 , 0.230	Depositor DCC
R_{free} test set	6394 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å ²)	17.0	Xtrriage
Anisotropy	0.427	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 43.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	10647	wwPDB-VP
Average B, all atoms (Å ²)	22.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 75.19 % 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 1.3258e-06. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹Intensities estimated from amplitudes.

²Theoretical 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: MG, NA, EPE

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	0.52	0/2487	0.87	14/3386 (0.4%)
1	B	0.50	0/2502	0.89	18/3407 (0.5%)
1	C	0.51	0/2500	0.87	18/3404 (0.5%)
1	D	0.51	0/2502	0.86	15/3407 (0.4%)
All	All	0.51	0/9991	0.87	65/13604 (0.5%)

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	1

There are no bond length outliers.

The worst 5 of 65 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	313	ARG	NE-CZ-NH2	-7.84	116.38	120.30
1	B	292	ASP	CB-CG-OD2	7.70	125.23	118.30
1	B	246	ARG	NE-CZ-NH1	7.34	123.97	120.30
1	A	91	ASP	CB-CG-OD1	7.08	124.67	118.30
1	A	236	ASP	CB-CG-OD2	6.94	124.55	118.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	220	ASP	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	2413	0	2244	32	0
1	B	2428	0	2264	29	0
1	C	2421	0	2254	25	0
1	D	2428	0	2264	28	0
2	A	2	0	0	0	0
2	B	2	0	0	0	0
2	C	2	0	0	0	0
2	D	1	0	0	0	0
3	A	15	0	17	0	0
3	B	15	0	17	1	0
3	C	15	0	17	0	0
3	D	15	0	17	1	0
4	B	1	0	0	0	0
4	C	1	0	0	0	0
5	A	226	0	0	3	1
5	B	230	0	0	4	0
5	C	213	0	0	3	0
5	D	219	0	0	2	1
All	All	10647	0	9094	114	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 6.

The worst 5 of 114 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:D:211:HIS:HD2	1:D:233:TRP:CZ2	1.90	0.90
1:B:197:GLU:OE2	5:B:2157:HOH:O	1.95	0.85
1:D:211:HIS:HD2	1:D:233:TRP:CE2	2.00	0.80
1:A:295:THR:HG23	1:A:298:ASN:OD1	1.84	0.78
1:D:273:ASP:OD1	5:D:2206:HOH:O	2.01	0.78

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:A:2219:HOH:O	5:D:2198:HOH:O[2_546]	2.14	0.06

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	293/346 (85%)	287 (98%)	6 (2%)	0	100	100
1	B	295/346 (85%)	289 (98%)	6 (2%)	0	100	100
1	C	295/346 (85%)	289 (98%)	6 (2%)	0	100	100
1	D	295/346 (85%)	290 (98%)	5 (2%)	0	100	100
All	All	1178/1384 (85%)	1155 (98%)	23 (2%)	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	260/308 (84%)	245 (94%)	15 (6%)	20	27
1	B	262/308 (85%)	252 (96%)	10 (4%)	33	47
1	C	262/308 (85%)	249 (95%)	13 (5%)	24	34
1	D	262/308 (85%)	241 (92%)	21 (8%)	12	15
All	All	1046/1232 (85%)	987 (94%)	59 (6%)	21	29

5 of 59 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	95[A]	SER
1	D	265	LEU
1	C	265	LEU
1	D	241	ASN
1	D	135	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 27 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	214	HIS
1	C	241	ASN
1	D	226	GLN
1	C	226	GLN
1	C	310	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 monosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 13 ligands modelled in this entry, 9 are monoatomic - leaving 4 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
3	EPE	B	2004	-	15,15,15	0.84	1 (6%)	18,20,20	1.21	2 (11%)
3	EPE	A	2002	-	15,15,15	0.86	1 (6%)	18,20,20	1.51	3 (16%)
3	EPE	C	2001	-	15,15,15	0.65	0	18,20,20	1.39	2 (11%)
3	EPE	D	2003	-	15,15,15	0.78	1 (6%)	18,20,20	1.25	2 (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
3	EPE	B	2004	-	-	7/9/19/19	0/1/1/1
3	EPE	A	2002	-	-	4/9/19/19	0/1/1/1
3	EPE	C	2001	-	-	2/9/19/19	0/1/1/1
3	EPE	D	2003	-	-	3/9/19/19	0/1/1/1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	2004	EPE	C10-S	2.94	1.81	1.77
3	A	2002	EPE	C10-S	2.86	1.81	1.77
3	D	2003	EPE	C10-S	2.37	1.80	1.77

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	2002	EPE	C6-N1-C2	3.24	116.12	108.83
3	C	2001	EPE	C6-C5-N4	3.22	117.25	110.64
3	D	2003	EPE	O2S-S-C10	2.99	110.52	106.92
3	B	2004	EPE	O2S-S-C10	2.93	110.44	106.92
3	C	2001	EPE	C5-N4-C3	2.80	115.12	108.83

There are no chirality outliers.

5 of 16 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	2002	EPE	S-C10-C9-N1
3	B	2004	EPE	S-C10-C9-N1
3	B	2004	EPE	C9-C10-S-O1S
3	B	2004	EPE	C9-C10-S-O2S

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
3	B	2004	EPE	C9-C10-S-O3S

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	2004	EPE	1	0
3	D	2003	EPE	1	0

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 [i](#)

6.1 Protein, DNA and RNA chains [i](#)

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, 95th 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	295/346 (85%)	-0.24	4 (1%) 75 80	14, 20, 29, 38	16 (5%)
1	B	297/346 (85%)	-0.25	1 (0%) 94 96	15, 21, 32, 36	21 (7%)
1	C	296/346 (85%)	-0.17	2 (0%) 87 91	15, 21, 32, 38	15 (5%)
1	D	297/346 (85%)	-0.20	5 (1%) 70 76	15, 21, 33, 40	20 (6%)
All	All	1185/1384 (85%)	-0.22	12 (1%) 82 86	14, 21, 32, 40	72 (6%)

The worst 5 of 12 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	71	PRO	3.9
1	D	354	LEU	3.6
1	B	71	PRO	3.3
1	D	70	GLY	2.8
1	A	221	PRO	2.6

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 monosaccharides 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, 95th 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(\AA^2)	Q<0.9
3	EPE	D	2003	15/15	0.72	0.26	37,39,45,46	6
3	EPE	B	2004	15/15	0.78	0.29	40,45,48,48	6
4	MG	B	1001	1/1	0.84	0.13	33,33,33,33	0
2	NA	D	1003	1/1	0.90	0.08	21,21,21,21	0
3	EPE	C	2001	15/15	0.91	0.18	28,35,37,37	6
3	EPE	A	2002	15/15	0.91	0.19	36,37,38,39	6
2	NA	A	1007	1/1	0.91	0.19	38,38,38,38	0
2	NA	B	1008	1/1	0.93	0.14	45,45,45,45	0
4	MG	C	1002	1/1	0.94	0.20	40,40,40,40	0
2	NA	A	1004	1/1	0.96	0.07	23,23,23,23	0
2	NA	C	1006	1/1	0.98	0.06	21,21,21,21	0
2	NA	C	1009	1/1	0.98	0.08	18,18,18,18	0
2	NA	B	1005	1/1	0.98	0.04	25,25,25,25	0

6.5 Other polymers [i](#)

There are no such residues in this entry.