



Full wwPDB X-ray Structure Validation Report ⓘ

Mar 8, 2018 – 08:42 pm GMT

PDB ID : 5T5C
Title : A Novel domain in human EXOG converts apoptotic endonuclease to DNA-repair enzyme
Authors : Szymanski, M.R.; Yin, W.Y.
Deposited on : 2016-08-30
Resolution : 1.85 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

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
Xtriage (Phenix) : 1.13
EDS : trunk30967
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac : 5.8.0158
CCP4 : 7.0 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : trunk30967

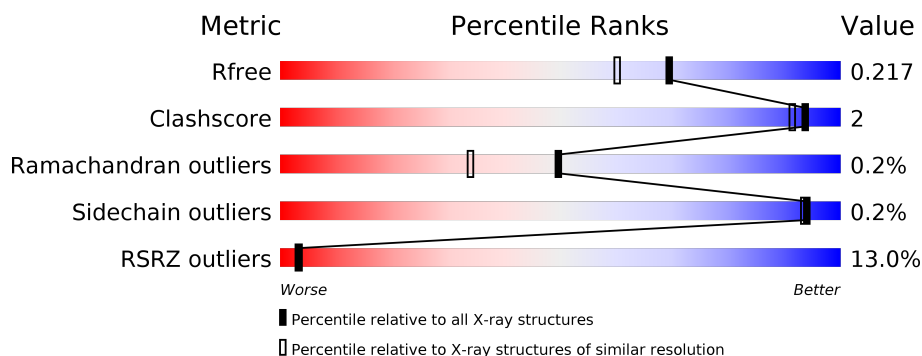
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.85 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}	111664	2111 (1.86-1.86)
Clashscore	122126	2258 (1.86-1.86)
Ramachandran outliers	120053	2234 (1.86-1.86)
Sidechain outliers	120020	2234 (1.86-1.86)
RSRZ outliers	108989	2075 (1.86-1.86)

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. 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	317	<div> <div>16%</div> <div>89%</div> <div>7%</div> </div>
1	B	317	<div> <div>7%</div> <div>92%</div> <div>7%</div> </div>
2	C	9	<div> <div>22%</div> <div>56%</div> <div>33%</div> <div>11%</div> </div>
2	E	9	<div> <div>33%</div> <div>56%</div> <div>33%</div> <div>11%</div> </div>
3	D	9	<div> <div>11%</div> <div>78%</div> <div>11%</div> <div>11%</div> </div>
3	F	9	<div> <div>33%</div> <div>67%</div> <div>22%</div> <div>11%</div> </div>

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 10712 atoms, of which 5030 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 Nuclease EXOG, mitochondrial.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	296	Total	C	H	N	O	S	0	0	0
			4711	1513	2335	404	448	11			
1	B	296	Total	C	H	N	O	S	0	0	0
			4711	1513	2335	404	448	11			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	58	MET	-	initiating methionine	UNP Q9Y2C4
A	140	ALA	HIS	engineered mutation	UNP Q9Y2C4
A	369	HIS	-	expression tag	UNP Q9Y2C4
A	370	HIS	-	expression tag	UNP Q9Y2C4
A	371	HIS	-	expression tag	UNP Q9Y2C4
A	372	HIS	-	expression tag	UNP Q9Y2C4
A	373	HIS	-	expression tag	UNP Q9Y2C4
A	374	HIS	-	expression tag	UNP Q9Y2C4
B	58	MET	-	initiating methionine	UNP Q9Y2C4
B	140	ALA	HIS	engineered mutation	UNP Q9Y2C4
B	369	HIS	-	expression tag	UNP Q9Y2C4
B	370	HIS	-	expression tag	UNP Q9Y2C4
B	371	HIS	-	expression tag	UNP Q9Y2C4
B	372	HIS	-	expression tag	UNP Q9Y2C4
B	373	HIS	-	expression tag	UNP Q9Y2C4
B	374	HIS	-	expression tag	UNP Q9Y2C4

- Molecule 2 is a DNA chain called DNA (5'-D(P*CP*TP*GP*AP*CP*GP*TP*GP*C)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	C	8	Total	C	H	N	O	P	0	0	0
			256	78	90	30	50	8			
2	E	8	Total	C	H	N	O	P	0	0	0
			256	78	90	30	50	8			

- Molecule 3 is a DNA chain called DNA (5'-D(P*GP*CP*AP*CP*GP*TP*CP*AP*G)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	D	8	Total	C	H	N	O	P	0	0	0
			253	77	90	31	47	8			
3	F	8	Total	C	H	N	O	P	0	0	0
			253	77	90	31	47	8			

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

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

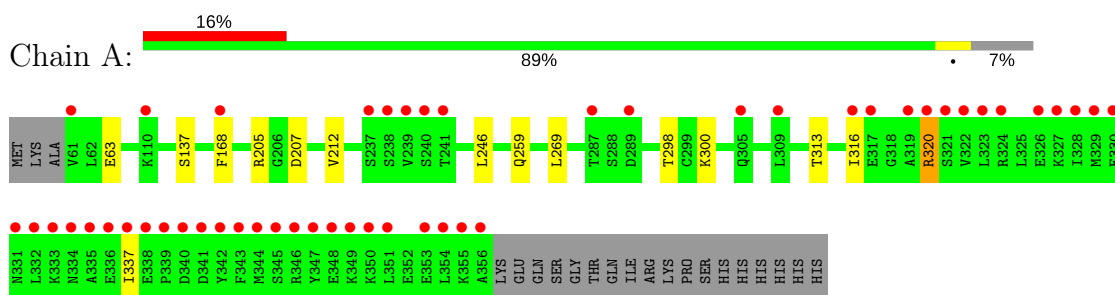
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	124	Total	O	0	0
			124	124		
5	B	134	Total	O	0	0
			134	134		
5	C	2	Total	O	0	0
			2	2		
5	D	2	Total	O	0	0
			2	2		
5	E	6	Total	O	0	0
			6	6		
5	F	2	Total	O	0	0
			2	2		

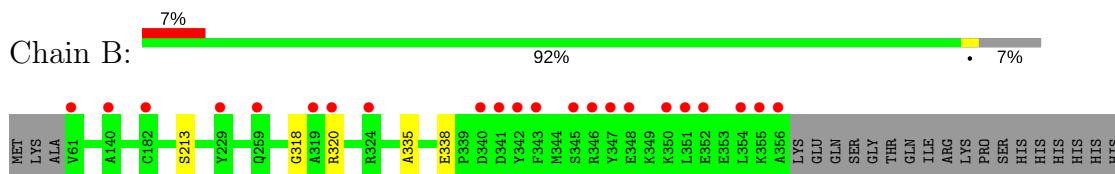
3 Residue-property plots [i](#)

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: Nuclease EXOG, mitochondrial



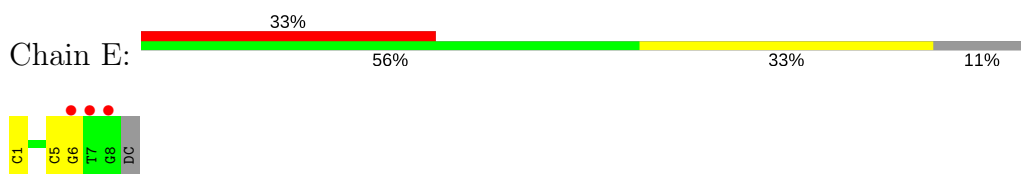
- Molecule 1: Nuclease EXOG, mitochondrial



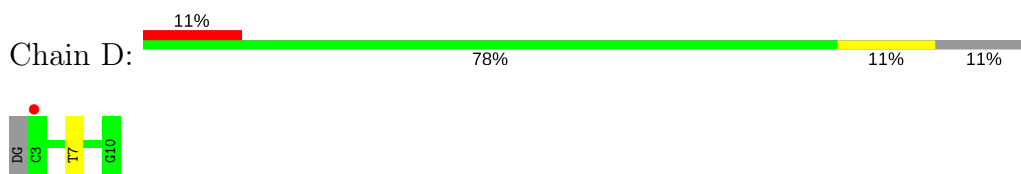
- Molecule 2: DNA (5'-D(P*CP*TP*GP*AP*CP*GP*TP*GP*C)-3')



- Molecule 2: DNA (5'-D(P*CP*TP*GP*AP*CP*GP*TP*GP*C)-3')

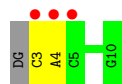


- Molecule 3: DNA (5'-D(P*GP*CP*AP*CP*GP*TP*CP*AP*G)-3')



- Molecule 3: DNA (5'-D(P*GP*CP*AP*CP*GP*TP*CP*AP*G)-3')

Chain F:



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	73.50Å 80.27Å 139.18Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.23 – 1.85 39.23 – 1.85	Depositor EDS
% Data completeness (in resolution range)	92.4 (39.23-1.85) 91.6 (39.23-1.85)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.07 (at 1.85Å)	Xtriage
Refinement program	PHENIX 1.10.1_2155	Depositor
R, R_{free}	0.186 , 0.216 0.186 , 0.217	Depositor DCC
R_{free} test set	7043 reflections (10.00%)	wwPDB-VP
Wilson B-factor (Å ²)	28.0	Xtriage
Anisotropy	0.679	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.41 , 54.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	10712	wwPDB-VP
Average B, all atoms (Å ²)	59.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.72% of the height of the origin peak. No significant pseudotranslation is detected.*

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

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	0.38	0/2432	0.51	0/3291
1	B	0.39	0/2432	0.53	0/3291
2	C	1.05	1/185 (0.5%)	0.90	0/282
2	E	1.10	1/185 (0.5%)	0.96	0/282
3	D	0.51	0/182	0.79	0/278
3	F	0.53	0/182	0.88	0/278
All	All	0.47	2/5598 (0.0%)	0.59	0/7702

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	1	DC	OP3-P	-10.38	1.48	1.61
2	E	1	DC	OP3-P	-9.86	1.49	1.61

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	2376	2335	2335	9	1
1	B	2376	2335	2335	3	1
2	C	166	90	91	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	E	166	90	91	3	0
3	D	163	90	90	1	0
3	F	163	90	90	1	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
5	A	124	0	0	0	0
5	B	134	0	0	0	0
5	C	2	0	0	0	0
5	D	2	0	0	0	0
5	E	6	0	0	0	0
5	F	2	0	0	0	0
All	All	5682	5030	5032	16	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 2.

All (16) 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:205:ARG:NH2	1:A:207:ASP:OD2	2.23	0.71
1:B:338:GLU:N	1:B:338:GLU:OE1	2.31	0.62
1:A:320:ARG:NH1	3:D:7:DT:OP2	2.32	0.60
1:A:313:THR:O	1:A:316:ILE:HG13	2.06	0.56
1:B:318:GLY:O	1:B:320:ARG:NH1	2.41	0.54
1:A:259:GLN:OE1	1:A:259:GLN:N	2.41	0.53
3:F:3:DC:H2''	3:F:4:DA:C8	2.44	0.53
2:E:5:DC:H2'	2:E:6:DG:N7	2.26	0.51
1:A:298:THR:HG23	1:A:300:LYS:H	1.76	0.50
1:A:212:VAL:O	1:B:213:SER:HA	2.14	0.48
1:A:63:GLU:OE1	1:A:63:GLU:N	2.48	0.45
1:A:246:LEU:HD21	1:A:269:LEU:HD22	2.00	0.43
1:A:137:SER:HB3	1:A:168:PHE:CD1	2.54	0.43
2:E:5:DC:C2'	2:E:6:DG:C8	3.02	0.43
2:E:5:DC:H2''	2:E:6:DG:C8	2.55	0.42
2:C:3:DG:H2'	2:C:4:DA:C8	2.55	0.41

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)
1:A:205:ARG:NH1	1:B:335:ALA:O[4_545]	2.19	0.01

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	294/317 (93%)	288 (98%)	5 (2%)	1 (0%)	43	27
1	B	294/317 (93%)	291 (99%)	3 (1%)	0	100	100
All	All	588/634 (93%)	579 (98%)	8 (1%)	1 (0%)	49	34

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	337	ILE

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	262/281 (93%)	261 (100%)	1 (0%)	92	90
1	B	262/281 (93%)	262 (100%)	0	100	100
All	All	524/562 (93%)	523 (100%)	1 (0%)	94	93

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	320	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

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 [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, 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	296/317 (93%)	1.06	50 (16%) 1 1	25, 44, 128, 175	0
1	B	296/317 (93%)	0.51	22 (7%) 14 15	22, 40, 84, 123	0
2	C	8/9 (88%)	1.25	2 (25%) 0 0	40, 52, 114, 148	0
2	E	8/9 (88%)	1.38	3 (37%) 0 0	37, 49, 120, 129	0
3	D	8/9 (88%)	1.38	1 (12%) 4 4	72, 90, 106, 134	0
3	F	8/9 (88%)	1.31	3 (37%) 0 0	54, 91, 108, 122	0
All	All	624/670 (93%)	0.82	81 (12%) 3 3	22, 42, 120, 175	0

All (81) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	323	LEU	9.3
1	A	319	ALA	9.2
1	A	328	ILE	8.1
1	A	331	ASN	8.0
1	A	334	ASN	7.9
1	A	332	LEU	7.8
1	A	341	ASP	7.6
1	A	239	VAL	6.9
1	A	337	ILE	6.7
1	B	342	TYR	6.7
1	A	316	ILE	6.6
1	A	348	GLU	6.6
1	A	342	TYR	6.5
1	A	322	VAL	6.5
1	A	340	ASP	6.5
1	A	345	SER	6.2
1	B	320	ARG	6.1
1	A	354	LEU	6.0
1	B	343	PHE	6.0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	241	THR	5.9
1	A	343	PHE	5.9
1	A	356	ALA	5.7
1	A	349	LYS	5.6
1	A	321	SER	5.4
2	C	8	DG	5.4
1	A	333	LYS	5.2
1	B	354	LEU	5.1
1	A	330	GLU	5.1
1	A	338	GLU	5.0
2	E	8	DG	5.0
1	B	346	ARG	4.7
1	A	320	ARG	4.7
1	A	355	LYS	4.6
1	A	168	PHE	4.6
1	A	350	LYS	4.5
1	A	240	SER	4.5
1	A	336	GLU	4.3
1	A	326	GLU	4.3
1	A	327	LYS	4.3
1	A	344	MET	4.3
1	A	289	ASP	4.2
1	A	339	PRO	4.0
1	A	335	ALA	4.0
1	A	309	LEU	3.8
3	D	3	DC	3.7
1	A	353	GLU	3.7
1	A	324	ARG	3.6
1	A	347	TYR	3.6
1	A	317	GLU	3.4
1	A	287	THR	3.3
1	B	355	LYS	3.3
2	E	7	DT	3.2
1	B	341	ASP	3.2
2	C	7	DT	3.0
1	A	351	LEU	3.0
1	B	351	LEU	2.9
1	B	319	ALA	2.7
1	A	346	ARG	2.7
1	B	356	ALA	2.6
1	B	347	TYR	2.6
2	E	6	DG	2.6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
3	F	3	DC	2.6
1	A	238	SER	2.4
1	B	324	ARG	2.4
3	F	5	DC	2.4
1	B	352	GLU	2.4
1	A	329	MET	2.3
1	B	229	TYR	2.3
1	A	305	GLN	2.3
1	A	237	SER	2.3
1	B	340	ASP	2.3
1	B	259	GLN	2.2
1	A	61	VAL	2.2
1	B	345	SER	2.2
1	B	350	LYS	2.1
3	F	4	DA	2.1
1	A	110	LYS	2.1
1	B	140	ALA	2.1
1	B	182	CYS	2.1
1	B	61	VAL	2.0
1	B	348	GLU	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, 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(Å ²)	Q<0.9
4	MG	A	401	1/1	0.97	0.14	33,33,33,33	0
4	MG	B	401	1/1	0.97	0.16	30,30,30,30	0

6.5 Other polymers [i](#)

There are no such residues in this entry.