



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 8, 2018 – 11:45 pm GMT

PDB ID : 2A07
Title : Crystal Structure of Foxp2 bound Specifically to DNA.
Authors : Stroud, J.C.; Wu, Y.; Bates, D.L.; Han, A.; Nowick, K.; Paabo, S.; Tong, H.;
Chen, L.
Deposited on : 2005-06-16
Resolution : 1.90 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 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

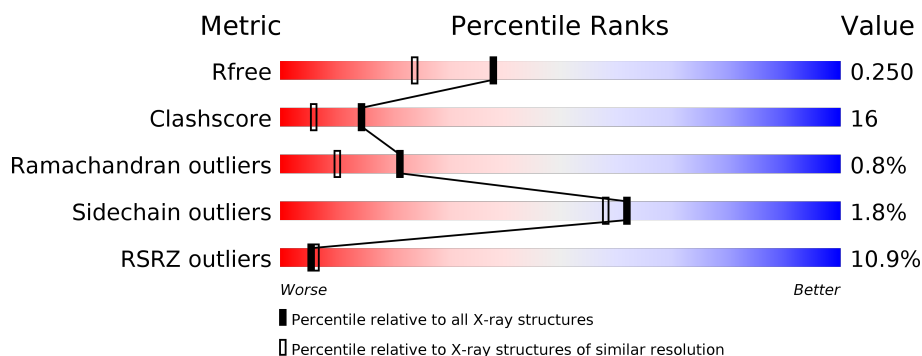
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.90 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	5502 (1.90-1.90)
Clashscore	122126	6115 (1.90-1.90)
Ramachandran outliers	120053	6048 (1.90-1.90)
Sidechain outliers	120020	6048 (1.90-1.90)
RSRZ outliers	108989	5379 (1.90-1.90)

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	21	
1	C	21	
2	B	21	
2	D	21	
3	F	93	
3	G	93	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	H	93	<div><div></div><div>15%</div><div>72%</div><div>11%</div><div>•</div><div>15%</div></div>
3	I	93	<div><div></div><div>18%</div><div>58%</div><div>27%</div><div>•</div><div>13%</div></div>
3	J	93	<div><div></div><div>2%</div><div>76%</div><div>12%</div><div>•</div><div>11%</div></div>
3	K	93	<div><div></div><div>3%</div><div>81%</div><div>9%</div><div>•</div><div>11%</div></div>

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 6446 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 DNA chain called 5'-D(*AP*AP*CP*TP*AP*TP*GP*AP*AP*AP*CP*AP*AP*AP*TP*TP*TP*TP*CP*CP*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	21	Total	C	N	O	P	0	0	0
			424	206	76	122	20			
1	C	21	Total	C	N	O	P	0	0	0
			424	206	76	122	20			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	21	Total	C	N	O	P	0	0	0
			431	209	76	126	20			
2	D	21	Total	C	N	O	P	0	0	0
			431	209	76	126	20			

- Molecule 3 is a protein called Forkhead box protein P2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	F	82	Total	C	N	O	S	0	0	0
			699	447	131	119	2			
3	G	79	Total	C	N	O	S	0	0	0
			674	431	125	116	2			
3	H	79	Total	C	N	O	S	0	0	0
			674	431	125	116	2			
3	I	81	Total	C	N	O	S	0	0	0
			688	441	127	118	2			
3	J	83	Total	C	N	O	S	0	0	0
			707	453	132	120	2			
3	K	83	Total	C	N	O	S	0	0	0
			707	453	132	120	2			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	502	ILE	ASP	ENGINEERED	UNP O15409
G	502	ILE	ASP	ENGINEERED	UNP O15409
H	502	ILE	ASP	ENGINEERED	UNP O15409
I	502	ILE	ASP	ENGINEERED	UNP O15409
J	502	ILE	ASP	ENGINEERED	UNP O15409
K	502	ILE	ASP	ENGINEERED	UNP O15409

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

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	G	1	Total Mg 1 1	0	0
4	J	1	Total Mg 1 1	0	0
4	K	1	Total Mg 1 1	0	0
4	H	1	Total Mg 1 1	0	0
4	I	1	Total Mg 1 1	0	0
4	F	1	Total Mg 1 1	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	52	Total O 52 52	0	0
5	B	55	Total O 55 55	0	0
5	C	44	Total O 44 44	0	0
5	D	59	Total O 59 59	0	0
5	F	39	Total O 39 39	0	0
5	G	39	Total O 39 39	0	0
5	H	54	Total O 54 54	0	0
5	I	45	Total O 45 45	0	0
5	J	106	Total O 106 106	0	0

Continued on next page...

Continued from previous page...

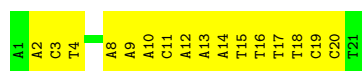
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	K	88	Total	O	0	0
			88	88		

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.

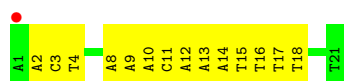
- Molecule 1: 5'-D(*AP*AP*CP*TP*AP*TP*GP*AP*AP*AP*CP*AP*AP*AP*TP*TP*TP*TP*CP*CP*T)-3'

Chain A: 

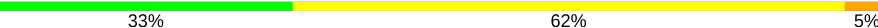


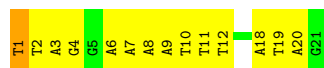
- Molecule 1: 5'-D(*AP*AP*CP*TP*AP*TP*GP*AP*AP*AP*CP*AP*AP*AP*TP*TP*TP*TP*CP*CP*T)-3'

Chain C: 



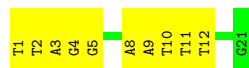
- Molecule 2: 5'-D(*TP*TP*AP*GP*GP*AP*AP*AP*AP*TP*TP*TP*GP*TP*TP*TP*CP*AP*TP*AP*G)-3'

Chain B: 



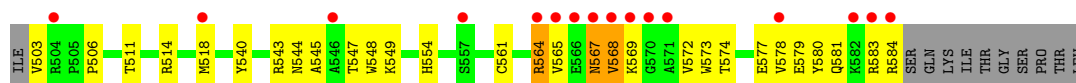
- Molecule 2: 5'-D(*TP*TP*AP*GP*GP*AP*AP*AP*AP*TP*TP*TP*GP*TP*TP*TP*CP*AP*TP*AP*G)-3'

Chain D: 

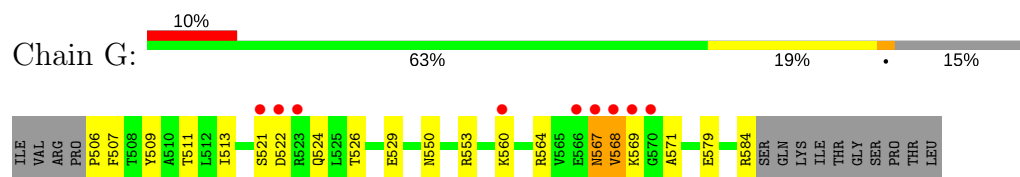


- Molecule 3: Forkhead box protein P2

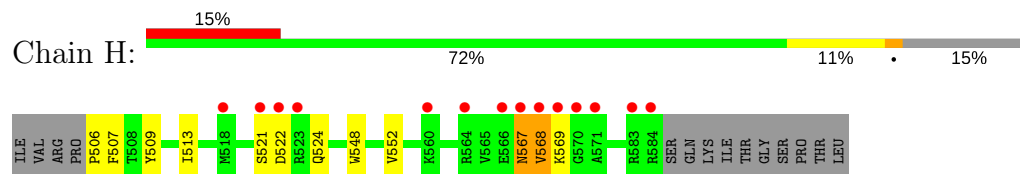
Chain F: 



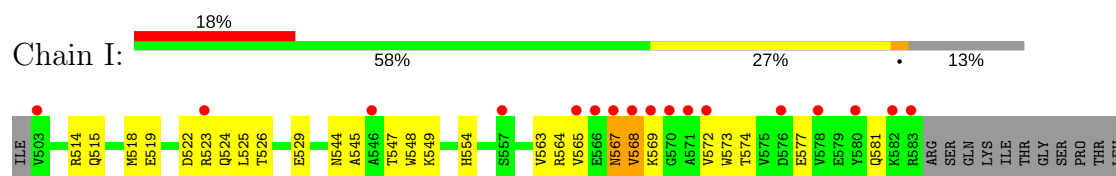
● Molecule 3: Forkhead box protein P2



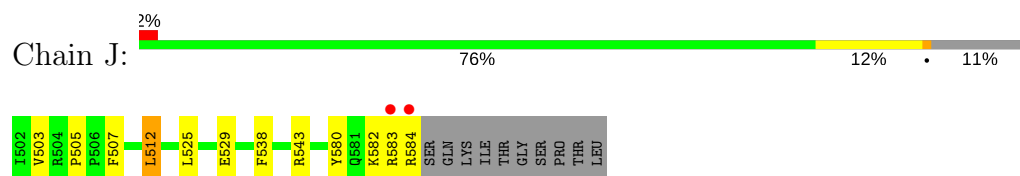
● Molecule 3: Forkhead box protein P2



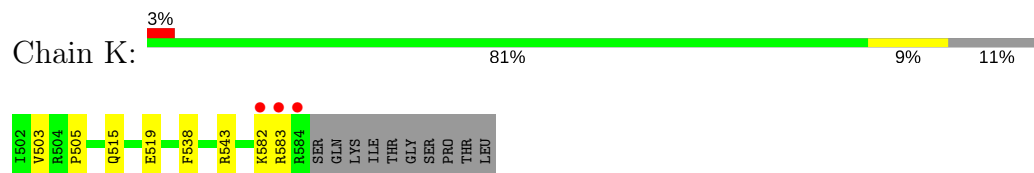
● Molecule 3: Forkhead box protein P2



● Molecule 3: Forkhead box protein P2



● Molecule 3: Forkhead box protein P2



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	67.54Å 124.21Å 67.67Å 90.00° 110.81° 90.00°	Depositor
Resolution (Å)	40.00 – 1.90 38.38 – 1.90	Depositor EDS
% Data completeness (in resolution range)	(Not available) (40.00-1.90) 96.3 (38.38-1.90)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.14 (at 1.89Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.222 , 0.244 0.231 , 0.250	Depositor DCC
R_{free} test set	8185 reflections (10.03%)	wwPDB-VP
Wilson B-factor (Å ²)	25.0	Xtriage
Anisotropy	0.178	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 43.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.486 for l,-k,h	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	6446	wwPDB-VP
Average B, all atoms (Å ²)	37.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.98% 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

5.1 Standard geometry

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	0.40	0/475	0.71	0/730
1	C	0.39	0/475	0.73	0/730
2	B	0.41	0/483	0.75	0/745
2	D	0.38	0/483	0.75	0/745
3	F	0.36	0/717	0.52	0/971
3	G	0.35	0/691	0.55	0/934
3	H	0.35	0/691	0.57	0/934
3	I	0.34	0/706	0.50	0/957
3	J	0.42	0/725	0.60	0/982
3	K	0.42	0/725	0.59	0/982
All	All	0.38	0/6171	0.62	0/8710

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
2	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
2	B	1	DT	Sidechain

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	424	0	240	27	0
1	C	424	0	240	24	0
2	B	431	0	242	23	0
2	D	431	0	242	17	0
3	F	699	0	693	31	0
3	G	674	0	665	17	0
3	H	674	0	665	11	0
3	I	688	0	680	21	0
3	J	707	0	704	7	0
3	K	707	0	704	4	0
4	F	1	0	0	0	0
4	G	1	0	0	0	0
4	H	1	0	0	0	0
4	I	1	0	0	0	0
4	J	1	0	0	0	0
4	K	1	0	0	0	0
5	A	52	0	0	0	0
5	B	55	0	0	2	0
5	C	44	0	0	0	0
5	D	59	0	0	1	0
5	F	39	0	0	2	0
5	G	39	0	0	1	0
5	H	54	0	0	1	0
5	I	45	0	0	0	0
5	J	106	0	0	1	0
5	K	88	0	0	1	0
All	All	6446	0	5075	180	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 16.

The worst 5 of 180 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)
2:B:1:DT:H2''	2:B:2:DT:H5''	1.36	1.06

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:G:568:VAL:HG23	3:G:569:LYS:H	1.34	0.93
3:F:568:VAL:HG13	3:F:569:LYS:H	1.35	0.91
1:A:2:DA:H2''	1:A:3:DC:H5'	1.52	0.91
1:C:12:DA:H2''	1:C:13:DA:H5''	1.55	0.89

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	
3	F	80/93 (86%)	75 (94%)	4 (5%)	1 (1%)	13	4
3	G	77/93 (83%)	75 (97%)	1 (1%)	1 (1%)	13	4
3	H	77/93 (83%)	72 (94%)	4 (5%)	1 (1%)	13	4
3	I	79/93 (85%)	73 (92%)	5 (6%)	1 (1%)	13	4
3	J	81/93 (87%)	79 (98%)	2 (2%)	0	100	100
3	K	81/93 (87%)	80 (99%)	1 (1%)	0	100	100
All	All	475/558 (85%)	454 (96%)	17 (4%)	4 (1%)	21	10

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	F	568	VAL
3	G	568	VAL
3	H	568	VAL
3	I	568	VAL

5.3.2 Protein sidechains ⓘ

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	
3	F	74/84 (88%)	72 (97%)	2 (3%)	48	40
3	G	71/84 (84%)	70 (99%)	1 (1%)	69	67
3	H	71/84 (84%)	70 (99%)	1 (1%)	69	67
3	I	73/84 (87%)	72 (99%)	1 (1%)	69	67
3	J	75/84 (89%)	73 (97%)	2 (3%)	48	40
3	K	75/84 (89%)	74 (99%)	1 (1%)	71	70
All	All	439/504 (87%)	431 (98%)	8 (2%)	62	58

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	H	567	ASN
3	K	543	ARG
3	J	512	LEU
3	G	567	ASN
3	I	567	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
3	H	544	ASN
3	H	567	ASN
3	K	524	GLN
3	H	515	GLN
3	J	555	ASN

5.3.3 RNA ⓘ

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 6 ligands modelled in this entry, 6 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 [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	21/21 (100%)	-0.07	0 100 100	13, 32, 64, 65	0
1	C	21/21 (100%)	-0.12	1 (4%) 30 34	14, 32, 62, 64	0
2	B	21/21 (100%)	-0.13	0 100 100	17, 23, 62, 72	0
2	D	21/21 (100%)	-0.25	0 100 100	17, 23, 59, 67	0
3	F	82/93 (88%)	0.76	16 (19%) 1 1	18, 41, 74, 89	0
3	G	79/93 (84%)	0.36	9 (11%) 5 5	17, 31, 72, 82	0
3	H	79/93 (84%)	0.91	14 (17%) 1 1	17, 33, 100, 111	0
3	I	81/93 (87%)	1.29	17 (20%) 1 1	18, 49, 104, 111	0
3	J	83/93 (89%)	-0.40	2 (2%) 59 62	12, 20, 43, 80	0
3	K	83/93 (89%)	-0.27	3 (3%) 42 46	11, 20, 52, 104	0
All	All	571/642 (88%)	0.35	62 (10%) 5 6	11, 30, 85, 111	0

The worst 5 of 62 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	H	568	VAL	17.5
3	I	570	GLY	12.5
3	H	569	LYS	10.4
3	I	568	VAL	10.4
3	K	583	ARG	10.3

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(\AA^2)	Q<0.9
4	MG	I	603	1/1	0.43	0.30	82,82,82,82	0
4	MG	F	601	1/1	0.75	0.18	69,69,69,69	0
4	MG	G	604	1/1	0.88	0.08	34,34,34,34	0
4	MG	H	602	1/1	0.93	0.13	38,38,38,38	0
4	MG	K	606	1/1	0.98	0.04	21,21,21,21	0
4	MG	J	605	1/1	0.99	0.04	18,18,18,18	0

6.5 Other polymers [i](#)

There are no such residues in this entry.