



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 21, 2020 – 10:48 AM EST

PDB ID : 3A9T  
Title : X-ray Structure of Bacillus pallidus D-Arabinose Isomerase Complex with L-Fucitol  
Authors : Takeda, K.; Yoshida, H.; Izumori, K.; Kamitori, S.  
Deposited on : 2009-11-05  
Resolution : 2.61 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](mailto: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
Mogul	:	1.8.0 (224370), CSD as540be (2019)
Xtriage (Phenix)	:	1.13
EDS	:	2.4
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)	:	2.4

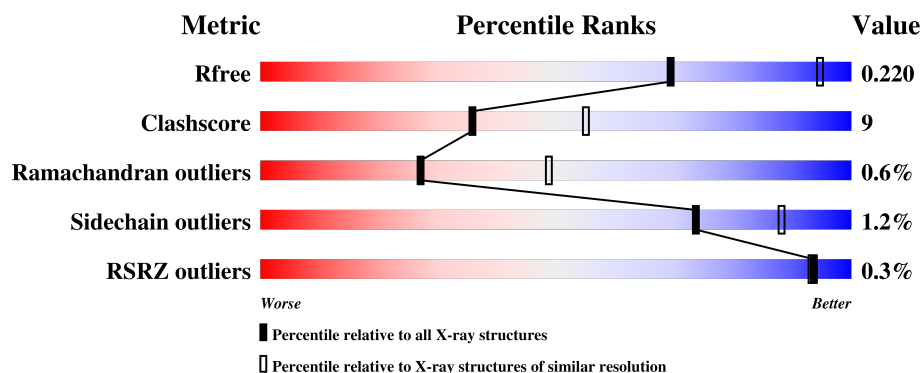
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.61 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	3285 (2.64-2.60)
Clashscore	122126	3641 (2.64-2.60)
Ramachandran outliers	120053	3586 (2.64-2.60)
Sidechain outliers	120020	3586 (2.64-2.60)
RSRZ outliers	108989	3218 (2.64-2.60)

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	595	<div> <div style="width: 78%;"></div> <div style="width: 20%;"></div> <div style="width: 2%;"></div> </div> <div>78% 20% .</div>
1	B	595	<div> <div style="width: 77%;"></div> <div style="width: 21%;"></div> <div style="width: 2%;"></div> </div> <div>77% 21% ..</div>
1	C	595	<div> <div style="width: 81%;"></div> <div style="width: 17%;"></div> <div style="width: 2%;"></div> </div> <div>81% 17% ..</div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 15018 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.

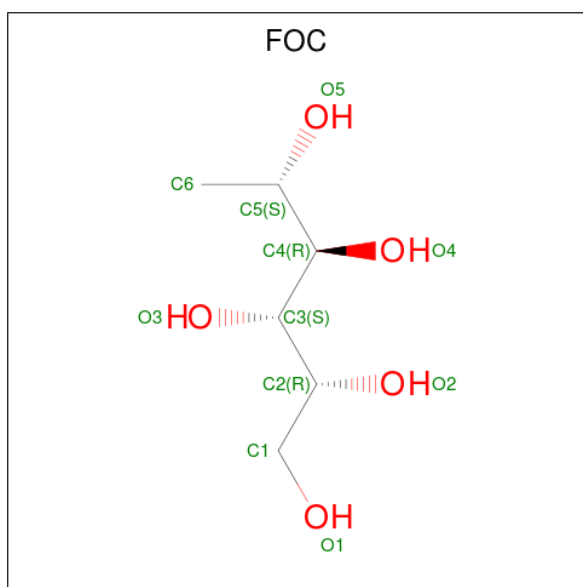
- Molecule 1 is a protein called D-arabinose isomerase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	589	Total	C	N	O	S	0	0	0
			4604	2922	779	877	26			
1	B	589	Total	C	N	O	S	0	0	0
			4604	2922	779	877	26			
1	C	589	Total	C	N	O	S	0	0	0
			4604	2922	779	877	26			

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	225	GLY	GLU	ENGINEERED	UNP C0SSE7
A	589	LYS	ASN	ENGINEERED	UNP C0SSE7
A	590	LEU	PHE	ENGINEERED	UNP C0SSE7
B	225	GLY	GLU	ENGINEERED	UNP C0SSE7
B	589	LYS	ASN	ENGINEERED	UNP C0SSE7
B	590	LEU	PHE	ENGINEERED	UNP C0SSE7
C	225	GLY	GLU	ENGINEERED	UNP C0SSE7
C	589	LYS	ASN	ENGINEERED	UNP C0SSE7
C	590	LEU	PHE	ENGINEERED	UNP C0SSE7

- Molecule 2 is FUCITOL (three-letter code: FOC) (formula: C<sub>6</sub>H<sub>14</sub>O<sub>5</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			11	6	5		
2	B	1	Total	C	O	0	0
			11	6	5		
2	C	1	Total	C	O	0	0
			11	6	5		

- Molecule 3 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	2	Total	Mn	0	0
			2	2		
3	A	2	Total	Mn	0	0
			2	2		
3	C	2	Total	Mn	0	0
			2	2		

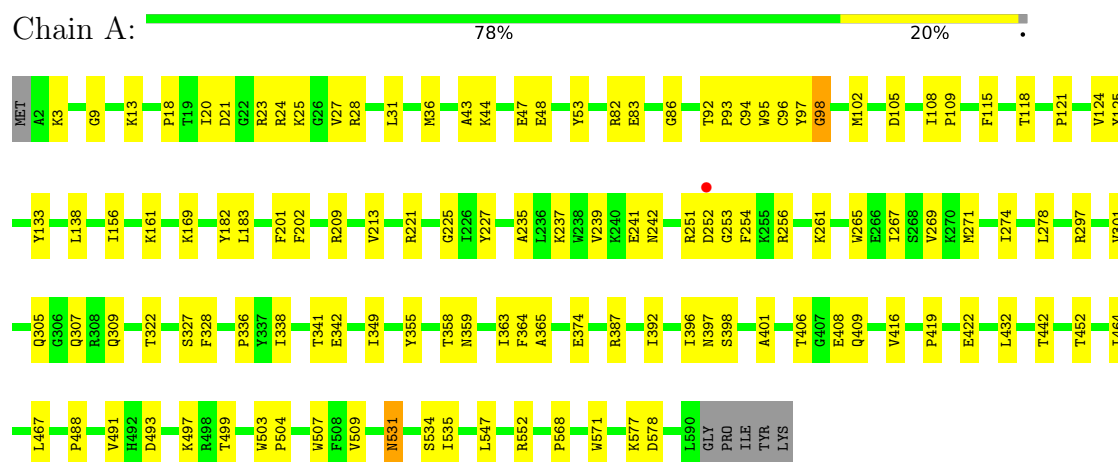
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	405	Total	O	0	0
			405	405		
4	B	403	Total	O	0	0
			403	403		
4	C	359	Total	O	0	0
			359	359		

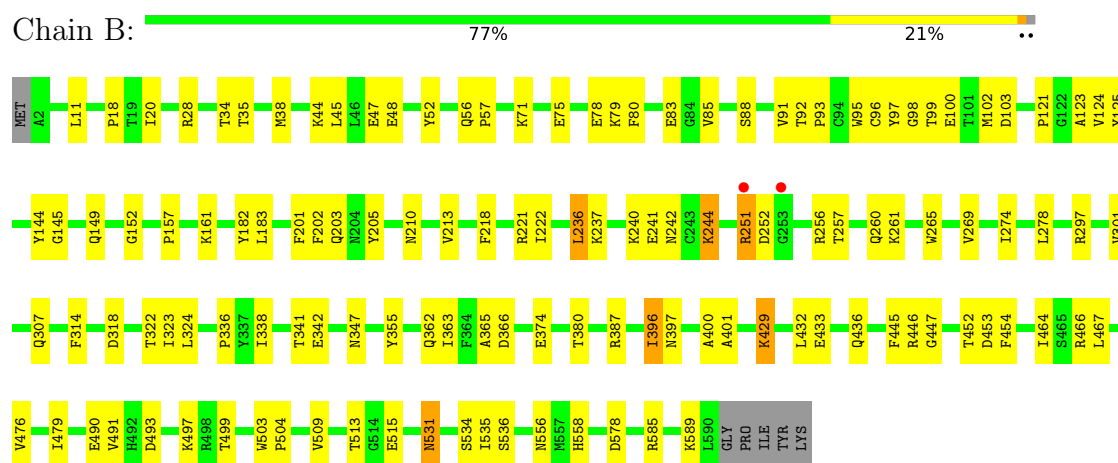
### 3 Residue-property plots

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.

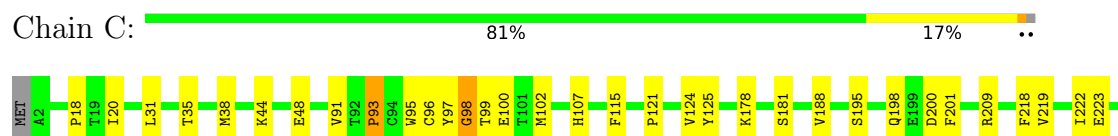
#### • Molecule 1: D-arabinose isomerase

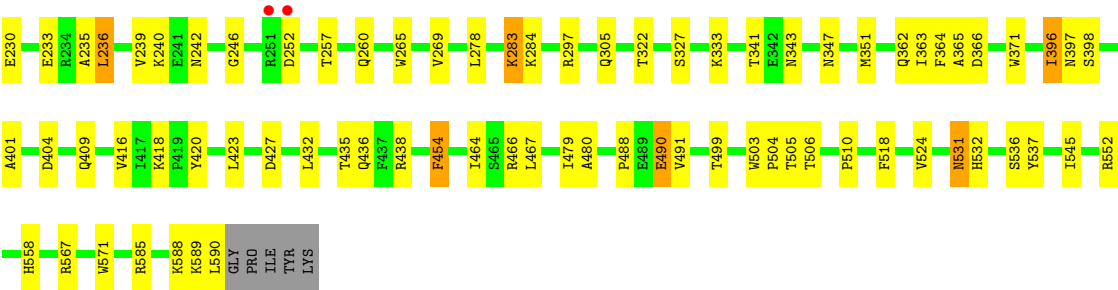


#### • Molecule 1: D-arabinose isomerase



#### • Molecule 1: D-arabinose isomerase





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	144.86Å 127.67Å 110.20Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.76 – 2.61 47.76 – 2.61	Depositor EDS
% Data completeness (in resolution range)	92.0 (47.76-2.61) 92.1 (47.76-2.61)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	6.04 (at 2.61Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.161 , 0.220 0.161 , 0.220	Depositor DCC
$R_{free}$ test set	5892 reflections (9.48%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	18.8	Xtriage
Anisotropy	0.860	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 51.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	15018	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	22.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.74% 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: MN, FOC

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	0.34	0/4706	0.57	0/6368
1	B	0.33	0/4706	0.57	1/6368 (0.0%)
1	C	0.34	0/4706	0.57	0/6368
All	All	0.34	0/14118	0.57	1/19104 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	103	ASP	N-CA-C	-5.01	97.48	111.00

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	4604	0	4515	80	0
1	B	4604	0	4515	85	0
1	C	4604	0	4515	78	0
2	A	11	0	13	0	0
2	B	11	0	13	0	0
2	C	11	0	13	0	0
3	A	2	0	0	0	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	2	0	0	0	0
3	C	2	0	0	0	0
4	A	405	0	0	9	0
4	B	403	0	0	10	0
4	C	359	0	0	8	0
All	All	15018	0	13584	239	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 9.

The worst 5 of 239 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:C:531:ASN:H	1:C:531:ASN:HD22	1.13	0.97
1:B:531:ASN:H	1:B:531:ASN:HD22	1.05	0.94
1:A:531:ASN:HD22	1:A:531:ASN:H	1.01	0.92
1:C:283:LYS:HD3	1:C:283:LYS:H	1.37	0.88
1:B:396:ILE:HD11	1:B:503:TRP:HB2	1.57	0.87

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	587/595 (99%)	556 (95%)	28 (5%)	3 (0%)	31	54
1	B	587/595 (99%)	557 (95%)	26 (4%)	4 (1%)	24	44
1	C	587/595 (99%)	559 (95%)	24 (4%)	4 (1%)	24	44
All	All	1761/1785 (99%)	1672 (95%)	78 (4%)	11 (1%)	27	49

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	252	ASP
1	C	252	ASP
1	A	98	GLY
1	A	398	SER
1	B	98	GLY

### 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	
1	A	482/487 (99%)	480 (100%)	2 (0%)	92	97
1	B	482/487 (99%)	475 (98%)	7 (2%)	67	84
1	C	482/487 (99%)	474 (98%)	8 (2%)	63	82
All	All	1446/1461 (99%)	1429 (99%)	17 (1%)	74	88

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	429	LYS
1	B	531	ASN
1	C	418	LYS
1	B	366	ASP
1	C	454	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 22 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	135	GLN
1	B	210	ASN
1	C	210	ASN
1	B	198	GLN
1	B	203	GLN

### 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 ⓘ

Of 9 ligands modelled in this entry, 6 are monoatomic - leaving 3 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	$\# Z  > 2$
2	FOC	A	601	3	9,10,10	0.26	0	10,13,13	0.76	1 (10%)
2	FOC	B	601	3	9,10,10	0.26	0	10,13,13	0.64	0
2	FOC	C	601	3	9,10,10	0.25	0	10,13,13	0.76	1 (10%)

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	FOC	A	601	3	-	1/14/14/14	-
2	FOC	B	601	3	-	0/14/14/14	-
2	FOC	C	601	3	-	0/14/14/14	-

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	601	FOC	C3-C4-C5	-2.18	110.18	113.21
2	C	601	FOC	C3-C4-C5	-2.03	110.39	113.21

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	601	FOC	O1-C1-C2-O2

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, 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	589/595 (98%)	-0.74	1 (0%) 94 95	9, 19, 34, 64	0
1	B	589/595 (98%)	-0.69	2 (0%) 93 93	8, 19, 34, 69	0
1	C	589/595 (98%)	-0.65	2 (0%) 93 93	7, 22, 39, 61	0
All	All	1767/1785 (98%)	-0.69	5 (0%) 93 93	7, 20, 36, 69	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	251	ARG	3.6
1	B	253	GLY	3.1
1	A	252	ASP	2.5
1	C	251	ARG	2.2
1	C	252	ASP	2.2

### 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, 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	B-factors( $\text{\AA}^2$ )	Q<0.9
3	MN	A	701	1/1	0.93	0.08	55,55,55,55	0
3	MN	B	701	1/1	0.94	0.07	64,64,64,64	0
2	FOC	A	601	11/11	0.94	0.20	30,32,37,39	0
2	FOC	B	601	11/11	0.94	0.24	26,29,33,34	0
2	FOC	C	601	11/11	0.95	0.18	25,27,29,30	0
3	MN	B	801	1/1	0.96	0.06	46,46,46,46	0
3	MN	C	801	1/1	0.96	0.08	53,53,53,53	0
3	MN	A	801	1/1	0.98	0.12	49,49,49,49	0
3	MN	C	701	1/1	0.98	0.07	45,45,45,45	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.