Full wwPDB X-ray Structure Validation Report (i)

Mar 9, 2024 – 05:53 PM EST

PDB ID	:	3OSF
Title	:	The structure of protozoan parasite Trichomonas vaginalis Myb2 in complex with MRE-2f-13 DNA
Authors	:	Jiang, I.; Tsai, C.K.; Chen, S.C.; Wang, S.H.; Amiraslanov, I.; Chang, C.F.; Wu, W.J.; Tai, J.H.; Liaw, Y.C.; Huang, T.H.
Deposited on	:	2010-09-09
Resolution	:	2.03 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 (i) 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 (i)) were used in the production of this report:

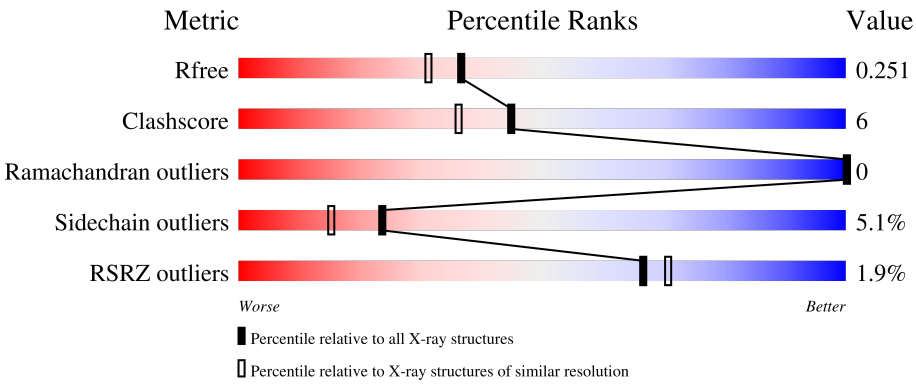
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.03 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	1692 (2.04-2.04)
Clashscore	141614	1773 (2.04-2.04)
Ramachandran outliers	138981	1752 (2.04-2.04)
Sidechain outliers	138945	1752 (2.04-2.04)
RSRZ outliers	127900	1672 (2.04-2.04)

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 <=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	126	2%, 68%, 12%, •, 18%
1	D	126	2%, 67%, 13%, •, 18%
2	B	13	54%, 31%, 15%
2	E	13	62%, 38%
3	C	13	77%, 23%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	13	69% 31%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 3130 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 MYB21.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	103	856	546	159	148	3	0	0	0
1	D	103	856	546	159	148	3	0	0	0

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	39	MET	-	expression tag	UNP Q58HP3
A	157	LEU	-	expression tag	UNP Q58HP3
A	158	GLU	-	expression tag	UNP Q58HP3
A	159	HIS	-	expression tag	UNP Q58HP3
A	160	HIS	-	expression tag	UNP Q58HP3
A	161	HIS	-	expression tag	UNP Q58HP3
A	162	HIS	-	expression tag	UNP Q58HP3
A	163	HIS	-	expression tag	UNP Q58HP3
A	164	HIS	-	expression tag	UNP Q58HP3
D	39	MET	-	expression tag	UNP Q58HP3
D	157	LEU	-	expression tag	UNP Q58HP3
D	158	GLU	-	expression tag	UNP Q58HP3
D	159	HIS	-	expression tag	UNP Q58HP3
D	160	HIS	-	expression tag	UNP Q58HP3
D	161	HIS	-	expression tag	UNP Q58HP3
D	162	HIS	-	expression tag	UNP Q58HP3
D	163	HIS	-	expression tag	UNP Q58HP3
D	164	HIS	-	expression tag	UNP Q58HP3

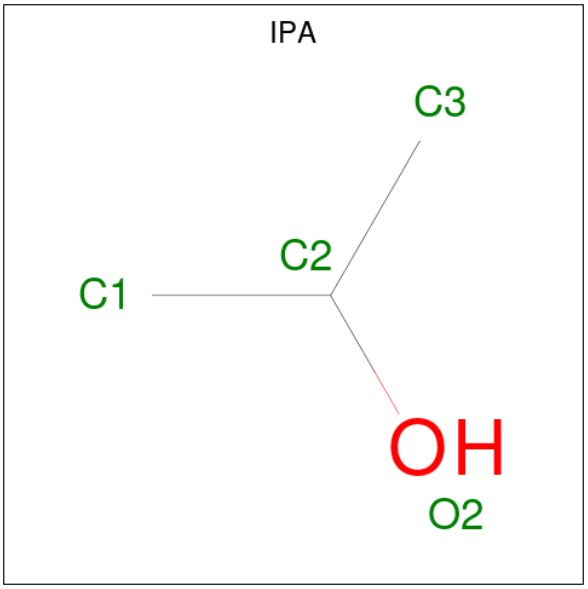
- Molecule 2 is a DNA chain called 5'-D(*CP*TP*GP*TP*AP*TP*CP*GP*TP*CP*TP*TP *G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	B	13	261	127	41	81	12	0	0	0
2	E	13	261	127	41	81	12	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	C	13	266	127	56	71	12	0	0	0
3	F	13	266	127	56	71	12	0	0	0

- Molecule 4 is ISOPROPYL ALCOHOL (three-letter code: IPA) (formula: C_3H_8O).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
4	A	1	4	3	1	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	92	92	92	0	0
5	B	29	29	29	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	C	38	38	38	0	0
			Total	O		
5	D	101	101	101	0	0
			Total	O		
5	E	47	47	47	0	0
			Total	O		
5	F	53	53	53	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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: MYB21

Chain A: 2% 68% 12% • 18%

MET VAL GLN VAL ASN LEU LYS ALA ALA K48 K51 L60 K72 R84 T101 Q116 F129 R132 T133 H136 I137 K138 W141 I144 S145 N146 K147 L148 G149 I150 PRO GLN THR GLN GLN MET LEU GLU HIS HIS HIS HIS HIS HIS

• Molecule 1: MYB21

Chain D: 2% 67% 13% • 18%

MET VAL GLN VAL ASN LEU LYS ALA ALA K48 K49 E55 L60 Q66 H67 K72 R84 H100 D108 L111 V112 Q113 K114 I115 Q116 F128 T133 H136 I137 K138 I144 L148 G149 I150 PRO GLN THR GLN GLN MET LEU GLU HIS HIS HIS HIS HIS HIS

HIS

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

Chain B: 54% 31% 15%

C1 T2 G3 T6 T11 T12 G13

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

Chain E: 62% 38%

C1 T2 G3 G8 T11 T12 G13

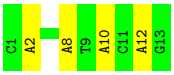
• Molecule 3: 5'-D(*CP*AP*AP*GP*AP*CP*GP*AP*TP*AP*CP*AP*G)-3'

Chain C: 77% 23%

C1 A8 T9 A10 G13

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

Chain F:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	73.94Å 77.35Å 84.41Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	36.65 – 2.03 37.05 – 2.03	Depositor EDS
% Data completeness (in resolution range)	98.7 (36.65-2.03) 98.7 (37.05-2.03)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.11 (at 2.03Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.6_289)	Depositor
R, R_{free}	0.206 , 0.258 0.200 , 0.251	Depositor DCC
R_{free} test set	1581 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	37.9	Xtriage
Anisotropy	0.150	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 49.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.014 for k,h,-l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	3130	wwPDB-VP
Average B, all atoms (Å^2)	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.48% of the height of the origin peak. No significant pseudotranslation is detected.*

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ 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: IPA

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	0.39	0/879	0.55	0/1187
1	D	0.40	0/879	0.56	0/1187
2	B	0.70	0/290	1.63	6/446 (1.3%)
2	E	0.84	0/290	1.62	4/446 (0.9%)
3	C	0.71	0/300	1.44	3/461 (0.7%)
3	F	0.74	0/300	1.46	3/461 (0.7%)
All	All	0.56	0/2938	1.10	16/4188 (0.4%)

There are no bond length outliers.

All (16) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	E	3	DG	O4'-C1'-N9	-13.56	98.51	108.00
2	B	6	DT	O4'-C1'-N1	-9.94	101.04	108.00
3	F	10	DA	O4'-C1'-N9	-9.62	101.27	108.00
2	B	11	DT	O4'-C1'-N1	-8.03	102.38	108.00
3	F	12	DA	O4'-C1'-N9	-6.78	103.26	108.00
2	B	12	DT	O4'-C1'-N1	-6.56	103.41	108.00
2	B	3	DG	O4'-C1'-N9	-6.54	103.42	108.00
3	F	8	DA	O4'-C1'-N9	-6.52	103.44	108.00
3	C	8	DA	O4'-C1'-N9	-6.50	103.45	108.00
3	C	1	DC	P-O3'-C3'	6.49	127.49	119.70
3	C	10	DA	O4'-C1'-N9	-6.10	103.73	108.00
2	B	6	DT	P-O3'-C3'	5.73	126.58	119.70
2	E	11	DT	O4'-C1'-N1	-5.49	104.16	108.00
2	E	8	DG	O4'-C1'-N9	-5.44	104.19	108.00
2	E	12	DT	O4'-C1'-N1	-5.11	104.42	108.00
2	B	2	DT	C5-C4-O4	-5.01	121.39	124.90

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	856	0	855	20	0
1	D	856	0	855	14	0
2	B	261	0	151	5	0
2	E	261	0	151	1	0
3	C	266	0	146	0	0
3	F	266	0	146	1	0
4	A	4	0	8	1	0
5	A	92	0	0	1	0
5	B	29	0	0	0	0
5	C	38	0	0	1	0
5	D	101	0	0	2	0
5	E	47	0	0	0	0
5	F	53	0	0	0	0
All	All	3130	0	2312	33	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 6.

All (33) 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:133:THR:H	1:A:136:HIS:CD2	2.02	0.78
1:A:116:GLN:HG3	1:D:148:LEU:HD13	1.68	0.74
1:D:66:GLN:NE2	1:D:67:HIS:CE1	2.58	0.71
1:A:133:THR:H	1:A:136:HIS:HD2	1.38	0.71
1:A:138:LYS:HE3	5:C:218:HOH:O	1.94	0.67
1:A:51:LYS:HG2	2:B:6:DT:H5'	1.78	0.66
1:A:147:LYS:HB2	1:D:116:GLN:HE21	1.64	0.63
1:D:133:THR:H	1:D:136:HIS:CD2	2.19	0.60
1:A:150:ILE:HG22	1:A:150:ILE:O	2.07	0.55
1:A:116:GLN:CG	1:D:148:LEU:HD13	2.36	0.53
1:D:133:THR:H	1:D:136:HIS:HD2	1.56	0.52

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:149:GLY:O	1:A:150:ILE:C	2.49	0.50
1:A:51:LYS:HA	2:B:6:DT:H5"	1.94	0.49
1:A:144:ILE:CG2	1:A:148:LEU:HD23	2.43	0.48
1:A:101:THR:HG23	4:A:1:IPA:O2	2.13	0.48
1:D:108:ASP:OD2	1:D:144:ILE:HD11	2.13	0.47
1:A:129:PHE:HB3	1:A:132:ARG:HG3	1.96	0.47
2:B:1:DC:H2'	2:B:2:DT:C6	2.51	0.46
1:D:55:GLU:HG2	5:D:197:HOH:O	2.15	0.46
1:D:138:LYS:HE2	1:D:138:LYS:HB3	1.75	0.45
2:E:1:DC:O5'	2:E:1:DC:H6	1.99	0.45
1:D:111:LEU:O	1:D:115:ILE:HG13	2.17	0.45
1:A:132:ARG:NH2	5:A:241:HOH:O	2.50	0.44
1:D:72:LYS:HE3	3:F:2:DA:OP1	2.17	0.44
1:A:51:LYS:HG2	2:B:6:DT:C5'	2.46	0.44
1:A:144:ILE:HG22	1:A:148:LEU:HD23	2.00	0.43
1:A:133:THR:N	1:A:136:HIS:HD2	2.11	0.42
1:A:141:TRP:HZ2	1:D:150:ILE:HD11	1.84	0.42
2:B:1:DC:H2"	2:B:2:DT:H5'	2.01	0.42
1:D:49:LYS:HG3	5:D:310:HOH:O	2.19	0.42
1:D:114:LYS:HE3	1:D:128:PHE:CG	2.55	0.42
1:A:148:LEU:HD13	1:A:148:LEU:HA	1.94	0.41
1:A:132:ARG:HA	1:A:136:HIS:HD2	1.86	0.40

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	101/126 (80%)	98 (97%)	3 (3%)	0	100	100
1	D	101/126 (80%)	98 (97%)	3 (3%)	0	100	100
All	All	202/252 (80%)	196 (97%)	6 (3%)	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	89/110 (81%)	84 (94%)	5 (6%)	21	12
1	D	89/110 (81%)	85 (96%)	4 (4%)	27	20
All	All	178/220 (81%)	169 (95%)	9 (5%)	24	15

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	60	LEU
1	A	84	ARG
1	A	132	ARG
1	A	146	ASN
1	A	148	LEU
1	D	49	LYS
1	D	60	LEU
1	D	84	ARG
1	D	148	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (8) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	66	GLN
1	A	92	ASN
1	A	136	HIS
1	D	66	GLN
1	D	67	HIS
1	D	100	HIS
1	D	116	GLN
1	D	136	HIS

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)

1 ligand is 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	$\#\|Z\| > 2$
4	IPA	A	1	-	3,3,3	0.47	0	3,3,3	0.40	0

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.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	1	IPA	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	103/126 (81%)	0.19	2 (1%)	66	71	27, 36, 53, 64	0
1	D	103/126 (81%)	0.18	3 (2%)	51	56	22, 37, 54, 64	0
2	B	13/13 (100%)	-0.32	0	100	100	31, 38, 46, 48	0
2	E	13/13 (100%)	-0.53	0	100	100	27, 35, 51, 55	0
3	C	13/13 (100%)	-0.57	0	100	100	31, 40, 49, 59	0
3	F	13/13 (100%)	-0.61	0	100	100	27, 31, 44, 47	0
All	All	258/304 (84%)	0.05	5 (1%)	66	71	22, 37, 54, 64	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	150	ILE	4.2
1	D	100	HIS	3.1
1	A	150	ILE	3.0
1	A	72	LYS	2.2
1	D	113	GLN	2.1

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(Å^2)	Q<0.9
4	IPA	A	1	4/4	0.95	0.15	38,44,44,44	0

6.5 Other polymers (i)

There are no such residues in this entry.