wwPDB X-ray Structure Validation Summary Report (i)

Jul 26, 2023 – 12:40 AM EDT

PDB ID	:	1AO7
Title	:	COMPLEX BETWEEN HUMAN T-CELL RECEPTOR, VIRAL PEPTIDE (TAX), AND HLA-A 0201
Authors	:	Garboczi, D.N.; Ghosh, P.; Utz, U.; Fan, Q.R.; Biddison, W.E.; Wiley, D.C.
Deposited on	:	1997-07-21
Resolution	:	2.60 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 (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:

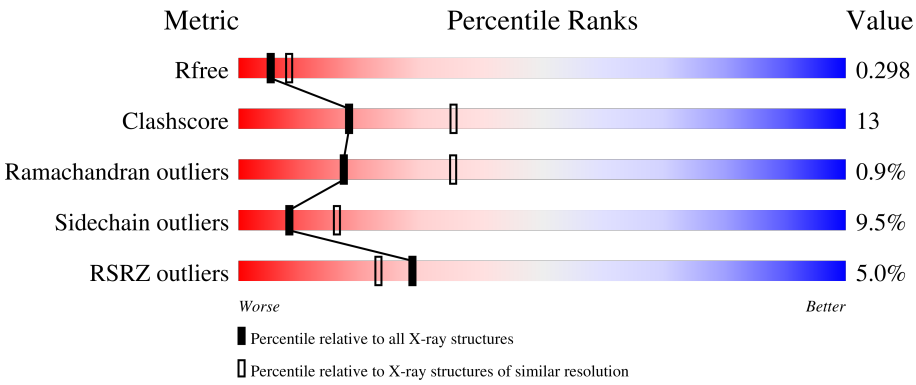
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.34
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.34

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 2.60 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	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

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 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	275	8%, 71%, 26%, •
2	B	100	%, 77%, 20%, •
3	C	9	56%, 44%
4	D	204	%, 38%, 16%, •, 44%
5	E	245	4%, 52%, 29%, •, 15%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
6	EMC	B	100	-	-	X	-

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 5711 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 HLA-A 0201.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	274	2237	1398	408	422	9	122	0	0

- Molecule 2 is a protein called BETA-2 MICROGLOBULIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	100	828	524	140	158	6	5	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	67	CYS	TYR	conflict	UNP P61769
B	91	CYS	LYS	conflict	UNP P61769

- Molecule 3 is a protein called TAX PEPTIDE.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
3	C	9	77	56	9	12	0	0	0

- Molecule 4 is a protein called T CELL RECEPTOR ALPHA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	D	115	886	553	148	182	3	25	0	0

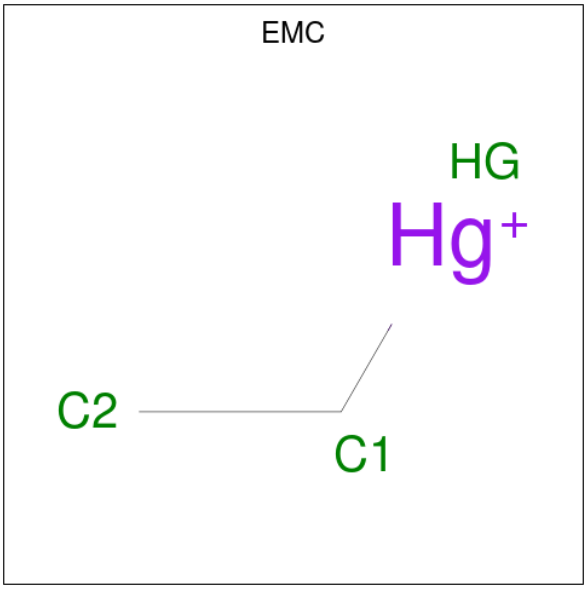
- Molecule 5 is a protein called T CELL RECEPTOR BETA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	E	209	1640	1041	286	305	8	40	0	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	95	ARG	SER	conflict	UNP 3002925
E	96	PRO	PHE	conflict	UNP 3002925
E	97	GLY	PRO	conflict	UNP 3002925
E	98	LEU	ARG	conflict	UNP 3002925
E	99	ALA	GLN	conflict	UNP 3002925
E	100	GLY	PRO	conflict	UNP 3002925
E	101	GLY	SER	conflict	UNP 3002925
E	102	ARG	TYR	conflict	UNP 3002925
E	103	PRO	ASN	conflict	UNP 3002925
E	107	TYR	PHE	conflict	UNP 3002925
E	116A	THR	LEU	conflict	UNP 3002925
E	191	ALA	CYS	conflict	UNP 3002925

- Molecule 6 is ETHYL MERCURY ION (three-letter code: EMC) (formula: C_2H_5Hg).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	Hg		
6	B	1	3	2	1	0	0
			Total	C	Hg		
6	B	1	3	2	1	0	0

- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
7	A	13	13	13	0	0

Continued on next page...

Continued from previous page...

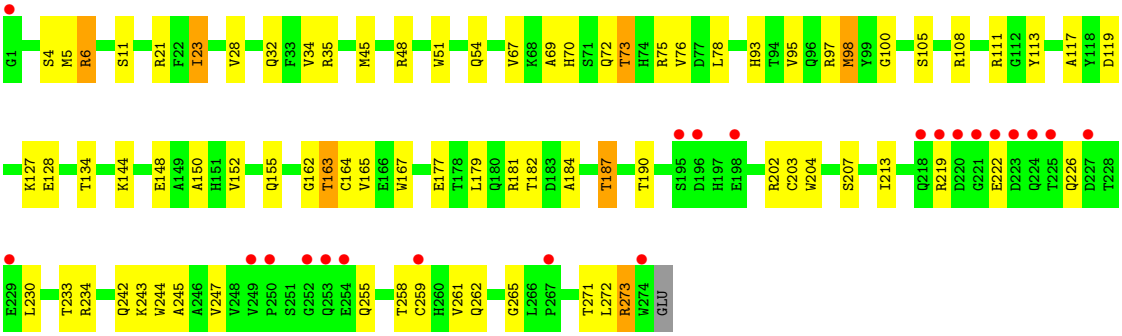
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
7	B	13	13	13	0	0
			Total	O		
7	C	1	1	1	0	0
			Total	O		
7	D	2	2	2	0	0
			Total	O		
7	E	8	8	8	0	0

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.

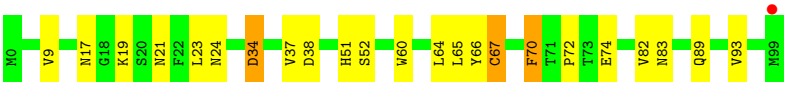
- Molecule 1: HLA-A 0201

Chain A:



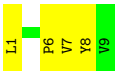
- Molecule 2: BETA-2 MICROGLOBULIN

Chain B:



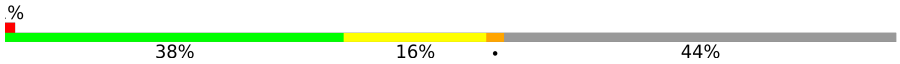
- Molecule 3: TAX PEPTIDE

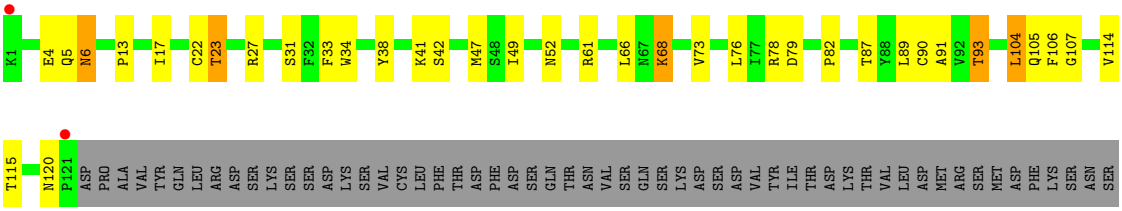
Chain C:



- Molecule 4: T CELL RECEPTOR ALPHA

Chain D:





ALA VAL ALA TRP SER ASN LYS SER ASP PHE ALA CYS ALA ASN ALA PHE ASN ASN SER ILE ILE PRO GLU ASP THR PHE PHE PRO SER PRO GLU SER SER

• Molecule 5: T CELL RECEPTOR BETA

Chain E: 4% 52% 29% • 15%

ASN ALA G3 Q6 T7 P8 V12 L13 S18 M19 T20 L21 Q25 Y31 M32 S33 W34 M41 L45 I46 H47 Q57 P61 V67 T71 L77 L80 P84 V89 Y90 F91 C92 R95 P96 G97 L98 A99 G100 G101 R102 P103 E105 Q106

T112 R113 L114 T115 V116 T116A E117 D118 L119 V122 F123 P124 P125 E126 V127 A128 V129 F130 GLU PRO SER GLU ALA GLU ILE SER HIS THR GLN LYS ALA THR L145 V146 C147 L148 F152 Y153 P154 D155 E158 L159 S160 W161 W162 V163 N164 G165 K166 E167 S170 T174 P178 L179

LYS GLU GLN PRO ALA LEU ASN ASP SER ARG Y190 A191 L192 R195 T201 F202 W203 Q204 N205 P206 R207 N208 H209 F210 V214 Q215 F216 Y217 G218 L219 SER GLU ASN ASP GLU TRP THR GLN ASP ARG A230 K231 P232 V233 T234 Q235 I236 V237 S238 A239 G243 R244 A245 D246

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	229.30Å 49.50Å 96.00Å 90.00° 89.60° 90.00°	Depositor
Resolution (Å)	6.00 – 2.60 11.95 – 2.34	Depositor EDS
% Data completeness (in resolution range)	94.9 (6.00-2.60) 85.9 (11.95-2.34)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	0.11	Depositor
$< I/\sigma(I) >$ [1]	1.81 (at 2.35Å)	Xtriage
Refinement program	X-PLOR 3.1	Depositor
R, R_{free}	0.245 , 0.320 0.232 , 0.298	Depositor DCC
R_{free} test set	3927 reflections (10.00%)	wwPDB-VP
Wilson B-factor (Å^2)	39.2	Xtriage
Anisotropy	0.189	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 77.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.022 for -h,-k,l	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	5711	wwPDB-VP
Average B, all atoms (Å^2)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.71% 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: EMC

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	0/2302	0.66	0/3125
2	B	0.40	0/850	0.75	1/1149 (0.1%)
3	C	0.51	0/80	0.82	0/108
4	D	0.41	0/905	0.74	1/1229 (0.1%)
5	E	0.39	0/1685	0.76	2/2292 (0.1%)
All	All	0.39	0/5822	0.72	4/7903 (0.1%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	64	LEU	CA-CB-CG	6.51	130.26	115.30
5	E	231	LYS	N-CA-C	6.18	127.69	111.00
5	E	13	LEU	CA-CB-CG	5.61	128.20	115.30
4	D	120	ASN	N-CA-C	5.51	125.87	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	2237	0	2090	45	0
2	B	828	0	789	15	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	C	77	0	79	4	0
4	D	886	0	845	27	0
5	E	1640	0	1570	59	0
6	B	6	0	0	2	0
7	A	13	0	0	0	0
7	B	13	0	0	0	0
7	C	1	0	0	0	0
7	D	2	0	0	0	0
7	E	8	0	0	0	0
All	All	5711	0	5373	142	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 13.

The worst 5 of 142 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:21:ARG:NE	1:A:23:ILE:HD11	1.91	0.86
1:A:187:THR:HG21	1:A:261:VAL:HG21	1.57	0.84
4:D:93:THR:HG21	4:D:104:LEU:HD22	1.60	0.83
1:A:32:GLN:HE21	1:A:48:ARG:HG3	1.44	0.81
5:E:21:LEU:HD12	5:E:77:LEU:HD23	1.62	0.80

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	272/275 (99%)	260 (96%)	12 (4%)	0	100	100
2	B	98/100 (98%)	95 (97%)	3 (3%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
3	C	7/9 (78%)	7 (100%)	0	0	100	100
4	D	113/204 (55%)	105 (93%)	7 (6%)	1 (1%)	17	35
5	E	201/245 (82%)	182 (90%)	14 (7%)	5 (2%)	5	9
All	All	691/833 (83%)	649 (94%)	36 (5%)	6 (1%)	17	35

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	D	6	ASN
5	E	119	LEU
5	E	231	LYS
5	E	164	ASN
5	E	155	ASP

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	230/231 (100%)	208 (90%)	22 (10%)	8	16
2	B	95/95 (100%)	89 (94%)	6 (6%)	18	36
3	C	8/8 (100%)	8 (100%)	0	100	100
4	D	99/182 (54%)	90 (91%)	9 (9%)	9	18
5	E	177/209 (85%)	156 (88%)	21 (12%)	5	9
All	All	609/725 (84%)	551 (90%)	58 (10%)	8	16

5 of 58 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	D	6	ASN
5	E	207	ARG
4	D	93	THR
5	E	195	ARG
5	E	118	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
5	E	208	ASN
5	E	37	GLN
4	D	5	GLN
5	E	11	GLN
2	B	83	ASN

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)

2 ligands are 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	$\#\|Z\| > 2$
6	EMC	B	101	2	1,2,2	0.70	0	-		
6	EMC	B	100	2	1,2,2	1.21	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 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	B	100	EMC	2	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	274/275 (99%)	-0.14	22 (8%)	12	9	6, 32, 119, 136	30 (10%)
2	B	100/100 (100%)	-0.67	1 (1%)	82	80	6, 25, 57, 81	2 (2%)
3	C	9/9 (100%)	-1.24	0	100	100	13, 19, 23, 24	0
4	D	115/204 (56%)	-0.45	2 (1%)	70	66	11, 33, 65, 76	7 (6%)
5	E	209/245 (85%)	-0.01	10 (4%)	30	24	16, 48, 90, 101	10 (4%)
All	All	707/833 (84%)	-0.24	35 (4%)	28	23	6, 35, 98, 136	49 (6%)

The worst 5 of 35 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	196	ASP	7.8
5	E	118	ASP	7.7
1	A	1	GLY	6.4
1	A	249	VAL	5.7
1	A	221	GLY	5.5

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
6	EMC	B	100	3/3	0.98	0.07	44,44,57,69	0
6	EMC	B	101	3/3	0.99	0.05	50,50,51,53	3

6.5 Other polymers (i)

There are no such residues in this entry.