Full wwPDB X-ray Structure Validation Report (i)

May 22, 2020 – 12:24 am BST

PDB ID	:	4H4F
Title	:	Crystal structure of human chymotrypsin C (CTRC) bound to inhibitor eglin c from Hirudo medicinalis
Authors	:	Batra, J.; Soares, A.S.; Radisky, E.S.
Deposited on	:	2012-09-17
Resolution	:	1.90 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 following versions of software and data (see references (1)) were used in the production of this report:

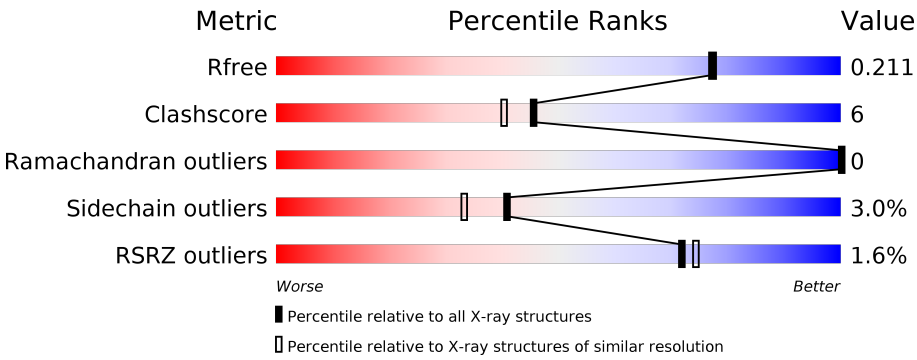
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 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}	130704	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (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 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	249	2% 81% 14% • •
2	B	70	71% 16% •• 10%
3	Q	10	100%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 2803 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 Chymotrypsin-C.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	239	1874	1184	335	344	11	0	0	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	245	HIS	-	EXPRESSION TAG	UNP Q99895
A	246	HIS	-	EXPRESSION TAG	UNP Q99895
A	247	HIS	-	EXPRESSION TAG	UNP Q99895
A	248	HIS	-	EXPRESSION TAG	UNP Q99895
A	249	HIS	-	EXPRESSION TAG	UNP Q99895
A	250	HIS	-	EXPRESSION TAG	UNP Q99895
A	251	HIS	-	EXPRESSION TAG	UNP Q99895
A	252	HIS	-	EXPRESSION TAG	UNP Q99895
A	253	HIS	-	EXPRESSION TAG	UNP Q99895
A	254	HIS	-	EXPRESSION TAG	UNP Q99895

- Molecule 2 is a protein called Eglin C.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
2	B	63	522	339	89	94	0	0	0

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1	MET	-	EXPRESSION TAG	UNP P01051
B	2	SER	-	EXPRESSION TAG	UNP P01051
B	3	MET	-	EXPRESSION TAG	UNP P01051
B	4	GLY	-	EXPRESSION TAG	UNP P01051
B	5	SER	-	EXPRESSION TAG	UNP P01051
B	6	GLU	-	EXPRESSION TAG	UNP P01051

Continued on next page...

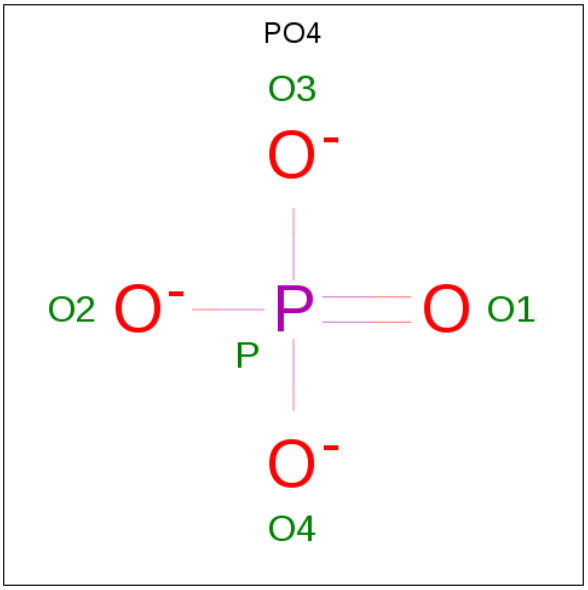
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	7	LEU	-	EXPRESSION TAG	UNP P01051

- Molecule 3 is a protein called Chymotrypsin-C.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	Q	10	72	47	11	13	1	0	0	0

- Molecule 4 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	O	P			
4	A	1	5	4	1		0	0
			Total	O	P			
4	A	1	5	4	1		0	0
			Total	O	P			
4	B	1	5	4	1		0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	251	251	251	0	0
			Total	O		
5	B	62	62	62	0	0

Continued on next page...

Continued from previous page...

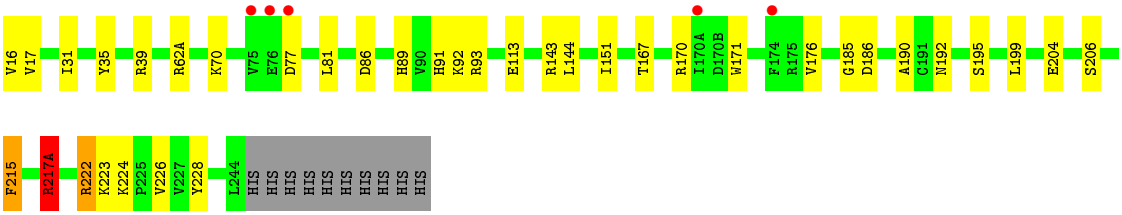
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	Q	7	7	7	0	0

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: Chymotrypsin-C

Chain A:

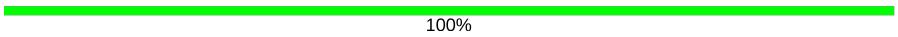


- Molecule 2: Eglin C

Chain B:

- Molecule 3: Chymotrypsin-C

Chain Q:



There are no outlier residues recorded for this chain.

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	56.27Å 76.25Å 81.82Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	55.78 – 1.90 38.13 – 1.90	Depositor EDS
% Data completeness (in resolution range)	97.3 (55.78-1.90) 97.4 (38.13-1.90)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	0.04	Depositor
$< I/\sigma(I) >$ [1]	6.51 (at 1.89Å)	Xtriage
Refinement program	REFMAC 5.5.0102	Depositor
R, R_{free}	0.157 , 0.210 0.158 , 0.211	Depositor DCC
R_{free} test set	1396 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å^2)	20.3	Xtriage
Anisotropy	1.284	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 53.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	2803	wwPDB-VP
Average B, all atoms (Å^2)	25.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.74% 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: PO4

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	1.32	6/1920 (0.3%)	1.14	12/2613 (0.5%)
2	B	1.26	2/540 (0.4%)	0.93	2/738 (0.3%)
3	Q	1.03	0/75	0.92	0/102
All	All	1.30	8/2535 (0.3%)	1.10	14/3453 (0.4%)

All (8) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	52	VAL	CB-CG1	6.47	1.66	1.52
1	A	204	GLU	CB-CG	-6.25	1.40	1.52
1	A	204	GLU	CD-OE1	5.77	1.32	1.25
2	B	12	GLU	CG-CD	5.57	1.60	1.51
1	A	17	VAL	CB-CG1	5.18	1.63	1.52
1	A	185	GLY	C-O	5.12	1.31	1.23
1	A	176	VAL	CB-CG1	5.01	1.63	1.52
1	A	215	PHE	CG-CD1	5.01	1.46	1.38

All (14) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	143	ARG	NE-CZ-NH2	-13.21	113.69	120.30
1	A	143	ARG	NE-CZ-NH1	12.05	126.33	120.30
1	A	186	ASP	CB-CG-OD1	7.76	125.28	118.30
1	A	144	LEU	O-C-N	-7.20	111.19	122.70
1	A	143	ARG	CG-CD-NE	-6.76	97.59	111.80
1	A	143	ARG	CD-NE-CZ	6.65	132.91	123.60
1	A	217(A)	ARG	NE-CZ-NH1	5.88	123.24	120.30
2	B	48	ARG	NE-CZ-NH2	-5.79	117.41	120.30
1	A	16	VAL	CA-CB-CG1	5.49	119.13	110.90
1	A	62(A)	ARG	NE-CZ-NH1	5.46	123.03	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	16	VAL	CG1-CB-CG2	5.40	119.53	110.90
1	A	86	ASP	CB-CG-OD1	5.31	123.08	118.30
2	B	48	ARG	NE-CZ-NH1	5.10	122.85	120.30
1	A	70	LYS	CD-CE-NZ	-5.00	100.19	111.70

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	1874	0	1829	20	1
2	B	522	0	498	13	0
3	Q	72	0	70	0	0
4	A	10	0	0	0	0
4	B	5	0	0	1	0
5	A	251	0	0	8	4
5	B	62	0	0	2	4
5	Q	7	0	0	0	0
All	All	2803	0	2397	31	5

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 (31) 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:31:GLN:HE21	2:B:31:GLN:H	0.96	0.93
5:A:473:HOH:O	2:B:65:HIS:HD2	1.58	0.86
2:B:31:GLN:N	2:B:31:GLN:HE21	1.78	0.82
2:B:31:GLN:NE2	2:B:31:GLN:H	1.80	0.71
1:A:91:HIS:HD2	1:A:93:ARG:H	1.38	0.70
1:A:93:ARG:NH2	5:A:562:HOH:O	2.26	0.69
2:B:48:ARG:NE	5:B:232:HOH:O	2.27	0.67
1:A:91:HIS:CD2	1:A:93:ARG:H	2.13	0.66

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:35:TYR:OH	4:B:101:PO4:O2	2.15	0.63
2:B:39:GLU:OE1	2:B:65:HIS:HE1	1.85	0.60
1:A:91:HIS:HE1	5:A:478:HOH:O	1.87	0.57
1:A:167:THR:HA	1:A:170:ARG:HD3	1.87	0.57
1:A:171:TRP:O	1:A:224:LYS:NZ	2.30	0.52
2:B:9:SER:OG	2:B:68:HIS:CE1	2.63	0.51
1:A:195:SER:OG	2:B:45:LEU:C	2.50	0.50
1:A:81:LEU:N	1:A:81:LEU:HD23	2.28	0.49
1:A:215:PHE:CD1	2:B:42:PRO:HB3	2.47	0.49
1:A:222:ARG:NH2	1:A:223:LYS:HE3	2.29	0.47
1:A:92:LYS:HD2	5:A:643:HOH:O	2.14	0.46
2:B:9:SER:HG	2:B:68:HIS:CE1	2.33	0.46
1:A:190:ALA:HB3	1:A:226:VAL:HG21	1.98	0.45
1:A:89:HIS:HD2	5:A:492:HOH:O	1.99	0.45
1:A:217(A):ARG:NH1	5:A:564:HOH:O	2.22	0.45
1:A:151:ILE:HD13	1:A:151:ILE:HG21	1.48	0.44
2:B:22:ARG:NH1	5:B:249:HOH:O	2.47	0.42
1:A:31:ILE:HD12	1:A:31:ILE:C	2.39	0.42
1:A:89:HIS:HE1	5:A:467:HOH:O	2.02	0.42
2:B:37:LEU:O	2:B:55:PHE:HA	2.19	0.42
1:A:199:LEU:HB2	1:A:228:TYR:CD2	2.56	0.41
1:A:223:LYS:NZ	5:A:531:HOH:O	2.51	0.41
2:B:9:SER:OG	2:B:68:HIS:ND1	2.49	0.41

All (5) 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:206:SER:CB	5:B:233:HOH:O[3_655]	1.53	0.67
5:A:544:HOH:O	5:B:262:HOH:O[2_665]	1.54	0.66
5:A:559:HOH:O	5:B:251:HOH:O[3_655]	1.70	0.50
5:A:431:HOH:O	5:A:557:HOH:O[3_645]	1.74	0.46
5:A:544:HOH:O	5:B:217:HOH:O[2_665]	2.15	0.05

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	237/249 (95%)	231 (98%)	6 (2%)	0	100	100
2	B	61/70 (87%)	60 (98%)	1 (2%)	0	100	100
3	Q	8/10 (80%)	8 (100%)	0	0	100	100
All	All	306/329 (93%)	299 (98%)	7 (2%)	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	203/213 (95%)	197 (97%)	6 (3%)	41	33
2	B	58/64 (91%)	56 (97%)	2 (3%)	37	28
3	Q	9/9 (100%)	9 (100%)	0	100	100
All	All	270/286 (94%)	262 (97%)	8 (3%)	41	33

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	39	ARG
1	A	77	ASP
1	A	113	GLU
1	A	192	ASN
1	A	217(A)	ARG
1	A	222	ARG
2	B	31	GLN
2	B	48	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (9) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	89	HIS
1	A	91	HIS
1	A	156	GLN
1	A	243	GLN
2	B	31	GLN
2	B	50	ASN
2	B	64	ASN
2	B	65	HIS
3	Q	9	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 carbohydrates in this entry.

5.6 Ligand geometry (i)

3 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	$\#\|Z\| > 2$
4	PO4	B	101	-	4,4,4	0.93	0	6,6,6	0.96	0
4	PO4	A	302	-	4,4,4	0.72	0	6,6,6	1.30	1 (16%)
4	PO4	A	301	-	4,4,4	0.96	0	6,6,6	1.27	2 (33%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	A	302	PO4	O3-P-O2	2.26	115.22	107.97
4	A	301	PO4	O4-P-O1	-2.18	102.91	110.89
4	A	301	PO4	O4-P-O2	2.06	114.58	107.97

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	B	101	PO4	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	239/249 (95%)	-0.09	5 (2%) 63 66	14, 21, 36, 55	0
2	B	63/70 (90%)	-0.15	0 100 100	18, 25, 35, 37	0
3	Q	10/10 (100%)	0.13	0 100 100	19, 21, 24, 26	0
All	All	312/329 (94%)	-0.10	5 (1%) 72 74	14, 22, 36, 55	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	76	GLU	3.4
1	A	77	ASP	2.7
1	A	170(A)	ILE	2.7
1	A	75	VAL	2.2
1	A	174	PHE	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	PO4	A	302	5/5	0.94	0.23	56,57,59,60	0
4	PO4	B	101	5/5	0.95	0.13	47,48,48,51	0
4	PO4	A	301	5/5	0.96	0.14	53,54,55,56	0

6.5 Other polymers (i)

There are no such residues in this entry.