



# wwPDB X-ray Structure Validation Summary Report

Feb 14, 2017 – 10:02 pm GMT

PDB ID : 3UTA  
Title : Crystal Structure of Nucleosome Core Particle Assembled with an Alpha-Satellite Sequence Containing Two TTAAA elements (NCP-TA2)  
Authors : Chua, E.Y.D.; Vasudevan, D.; Davey, G.E.; Wu, B.; Davey, C.A.  
Deposited on : 2011-11-25  
Resolution : 2.07 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

<http://wwpdb.org/validation/2016/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) : 1.9-1692  
EDS : trunk28620  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : recalc28949

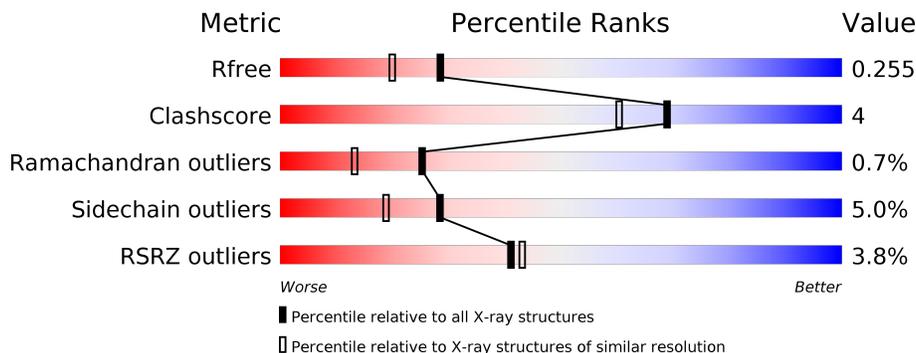
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.07 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))
$R_{free}$	100719	2028 (2.08-2.04)
Clashscore	112137	2143 (2.08-2.04)
Ramachandran outliers	110173	2126 (2.08-2.04)
Sidechain outliers	110143	2126 (2.08-2.04)
RSRZ outliers	101464	2035 (2.08-2.04)

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 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	135	<p>2% 63% 8% 28%</p>
1	E	135	<p>61% 7% 28%</p>
2	B	102	<p>4% 74% 5% 20%</p>
2	F	102	<p>3% 76% 6% 15%</p>
3	C	129	<p>3% 70% 10% 18%</p>
3	G	129	<p>7% 77% 5% 18%</p>

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
4	D	125	
4	H	125	
5	I	145	
6	J	145	

## 2 Entry composition [i](#)

There are 9 unique types of molecules in this entry. The entry contains 12155 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 Histone H3.2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	97	802	506	155	138	3	4	0	0
1	E	97	802	506	155	138	3	4	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	102	ALA	GLY	SEE REMARK 999	UNP P84233
E	102	ALA	GLY	SEE REMARK 999	UNP P84233

- Molecule 2 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	82	653	412	127	113	1	0	0	0
2	F	87	703	442	142	118	1	0	0	0

- Molecule 3 is a protein called Histone H2A.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
3	C	106	818	516	160	142	0	0	0
3	G	106	818	516	160	142	0	0	0

- Molecule 4 is a protein called Histone H2B 1.1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	D	95	745	469	134	140	2	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	H	95	745	469	134	140	2	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	29	THR	SER	SEE REMARK 999	UNP P02281
H	29	THR	SER	SEE REMARK 999	UNP P02281

- Molecule 5 is a DNA chain called 145-mer DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
5	I	145	2970	1424	535	867	144	0	0	0

- Molecule 6 is a DNA chain called 145-mer DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
6	J	145	2969	1424	532	869	144	0	0	0

- Molecule 7 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	G	1	Total Cl 1 1	0	0
7	A	1	Total Cl 1 1	0	0
7	C	1	Total Cl 1 1	0	0
7	E	1	Total Cl 1 1	0	0

- Molecule 8 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
8	J	7	Total Mn 7 7	0	0
8	I	8	Total Mn 8 8	0	0
8	F	1	Total Mn 1 1	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	E	1	Total 1	Mn 1	0	0

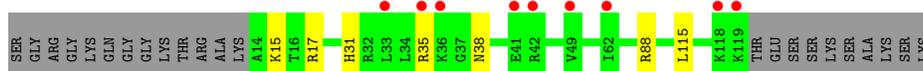
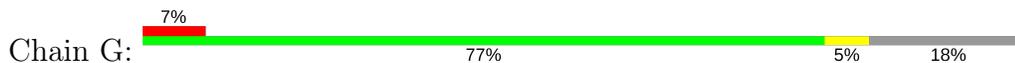
- Molecule 9 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	A	14	Total 14	O 14	0	0
9	B	15	Total 15	O 15	0	0
9	C	14	Total 14	O 14	0	0
9	D	5	Total 5	O 5	0	0
9	E	29	Total 29	O 29	0	0
9	F	16	Total 16	O 16	0	0
9	G	9	Total 9	O 9	0	0
9	H	5	Total 5	O 5	0	0
9	J	2	Total 2	O 2	0	0





● Molecule 3: Histone H2A



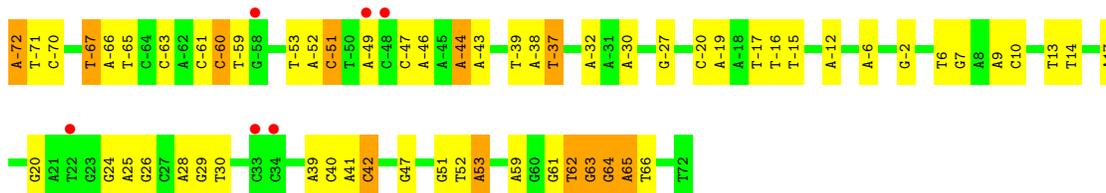
● Molecule 4: Histone H2B 1.1



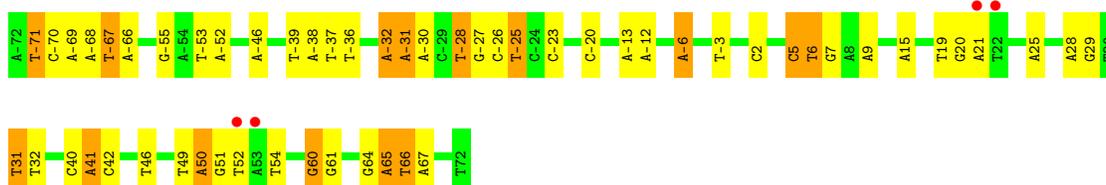
● Molecule 4: Histone H2B 1.1



● Molecule 5: 145-mer DNA



● Molecule 6: 145-mer DNA



## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	106.51Å 109.91Å 182.33Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	94.13 – 2.07 94.13 – 2.07	Depositor EDS
% Data completeness (in resolution range)	75.5 (94.13-2.07) 75.5 (94.13-2.07)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.06	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.18 (at 2.07Å)	Xtrriage
Refinement program	REFMAC 5.5.0109	Depositor
R, $R_{free}$	0.240 , 0.266 0.233 , 0.255	Depositor DCC
$R_{free}$ test set	2009 reflections (2.07%)	DCC
Wilson B-factor (Å <sup>2</sup> )	45.4	Xtrriage
Anisotropy	0.336	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 48.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.017 for k,h,-l	Xtrriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	12155	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	67.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.06% 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, CL

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	0.89	1/814 (0.1%)	0.62	1/1092 (0.1%)
1	E	1.07	1/814 (0.1%)	0.87	4/1092 (0.4%)
2	B	0.53	0/660	0.60	0/883
2	F	0.66	0/711	0.67	0/948
3	C	0.62	0/828	0.74	2/1117 (0.2%)
3	G	0.48	0/828	0.59	0/1117
4	D	0.57	0/756	0.61	0/1015
4	H	0.47	0/756	0.55	0/1015
5	I	0.75	0/3332	1.45	35/5141 (0.7%)
6	J	0.72	0/3330	1.44	45/5138 (0.9%)
All	All	0.72	2/12829 (0.0%)	1.16	87/18558 (0.5%)

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
1	A	0	1
1	E	0	1
All	All	0	2

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	134	ARG	CD-NE	-23.76	1.06	1.46
1	A	134	ARG	CD-NE	21.20	1.82	1.46

The worst 5 of 87 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	E	134	ARG	CD-NE-CZ	11.24	139.34	123.60
6	J	-25	DT	O4'-C1'-N1	9.11	114.37	108.00
6	J	-52	DA	P-O3'-C3'	8.16	129.49	119.70
3	C	81	ARG	NE-CZ-NH1	7.68	124.14	120.30
5	I	42	DC	O4'-C1'-N1	7.64	113.35	108.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	134	ARG	Sidechain
1	E	134	ARG	Sidechain

## 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	802	0	841	6	0
1	E	802	0	841	13	0
2	B	653	0	696	3	0
2	F	703	0	755	4	0
3	C	818	0	877	11	0
3	G	818	0	877	7	0
4	D	745	0	773	9	0
4	H	745	0	773	2	0
5	I	2970	0	1643	33	0
6	J	2969	0	1644	30	0
7	A	1	0	0	1	0
7	C	1	0	0	0	0
7	E	1	0	0	0	0
7	G	1	0	0	0	0
8	E	1	0	0	0	0
8	F	1	0	0	0	0
8	I	8	0	0	0	0
8	J	7	0	0	0	0
9	A	14	0	0	0	0
9	B	15	0	0	0	0
9	C	14	0	0	0	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
9	D	5	0	0	0	0
9	E	29	0	0	0	0
9	F	16	0	0	0	0
9	G	9	0	0	0	0
9	H	5	0	0	0	0
9	J	2	0	0	0	0
All	All	12155	0	9720	92	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 4.

The worst 5 of 92 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)
5:I:-44:DA:H2''	5:I:-43:DA:C8	2.06	0.90
2:F:17:ARG:O	2:F:18:HIS:HB2	1.77	0.82
5:I:-47:DC:H4'	5:I:-46:DA:H5'	1.62	0.82
2:F:20:LYS:HE2	2:F:21:VAL:H	1.50	0.74
6:J:65:DA:H2''	6:J:66:DT:OP2	1.91	0.70

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	95/135 (70%)	95 (100%)	0	0	100	100
1	E	95/135 (70%)	94 (99%)	1 (1%)	0	100	100
2	B	80/102 (78%)	77 (96%)	1 (1%)	2 (2%)	6	1
2	F	85/102 (83%)	82 (96%)	2 (2%)	1 (1%)	15	6
3	C	104/129 (81%)	101 (97%)	3 (3%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
3	G	104/129 (81%)	101 (97%)	3 (3%)	0	100	100
4	D	93/125 (74%)	92 (99%)	0	1 (1%)	17	6
4	H	93/125 (74%)	91 (98%)	1 (1%)	1 (1%)	17	6
All	All	749/982 (76%)	733 (98%)	11 (2%)	5 (1%)	25	13

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	D	101	GLY
2	F	18	HIS
4	H	101	GLY
2	B	22	LEU
2	B	23	ARG

### 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	85/110 (77%)	80 (94%)	5 (6%)	23	13
1	E	85/110 (77%)	82 (96%)	3 (4%)	41	33
2	B	67/78 (86%)	63 (94%)	4 (6%)	22	13
2	F	72/78 (92%)	66 (92%)	6 (8%)	13	6
3	C	84/101 (83%)	78 (93%)	6 (7%)	17	9
3	G	84/101 (83%)	82 (98%)	2 (2%)	54	49
4	D	81/105 (77%)	78 (96%)	3 (4%)	39	31
4	H	81/105 (77%)	78 (96%)	3 (4%)	39	31
All	All	639/788 (81%)	607 (95%)	32 (5%)	28	19

5 of 32 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	C	119	LYS

Continued on next page...

*Continued from previous page...*

Mol	Chain	Res	Type
4	D	31	LYS
4	H	31	LYS
4	D	30	ARG
1	E	49	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
4	D	79	HIS
3	G	31	HIS
4	D	92	GLN
3	C	31	HIS
4	D	106	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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 21 ligands modelled in this entry, 21 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, 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	97/135 (71%)	0.65	3 (3%) 49 53	29, 41, 57, 80	1 (1%)
1	E	97/135 (71%)	0.49	0 100 100	21, 34, 52, 67	1 (1%)
2	B	82/102 (80%)	0.57	4 (4%) 30 31	30, 39, 59, 83	0
2	F	87/102 (85%)	0.47	3 (3%) 46 48	22, 33, 52, 90	0
3	C	106/129 (82%)	0.70	4 (3%) 41 43	23, 37, 59, 73	0
3	G	106/129 (82%)	0.62	9 (8%) 11 11	31, 47, 65, 75	0
4	D	95/125 (76%)	0.72	4 (4%) 37 38	27, 38, 69, 82	0
4	H	95/125 (76%)	0.67	3 (3%) 48 51	34, 49, 76, 83	0
5	I	145/145 (100%)	0.33	6 (4%) 38 39	45, 91, 131, 147	0
6	J	145/145 (100%)	0.29	4 (2%) 53 57	45, 89, 132, 147	0
All	All	1055/1272 (82%)	0.53	40 (3%) 41 43	21, 45, 115, 147	2 (0%)

The worst 5 of 40 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	119	LYS	7.8
1	A	39	HIS	5.9
2	B	21	VAL	5.8
4	D	122	LYS	4.7
4	D	29	THR	3.9

### 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
8	MN	F	1016	1/1	0.07	0.26	1.57	156,156,156,156	0
8	MN	E	1001	1/1	1.00	0.17	0.81	30,30,30,30	0
8	MN	J	1004	1/1	0.97	0.10	-1.69	86,86,86,86	0
7	CL	G	2003	1/1	0.97	0.10	-1.90	51,51,51,51	0
7	CL	C	2004	1/1	0.98	0.10	-3.63	47,47,47,47	0
8	MN	I	1006	1/1	0.63	0.16	-	100,100,100,100	0
8	MN	I	1014	1/1	0.84	0.08	-	107,107,107,107	0
8	MN	I	1011	1/1	-0.01	0.14	-	130,130,130,130	0
8	MN	J	1013	1/1	0.76	0.10	-	111,111,111,111	0
8	MN	I	1003	1/1	0.70	0.09	-	108,108,108,108	0
8	MN	I	1010	1/1	0.96	0.20	-	72,72,72,72	0
8	MN	J	1005	1/1	0.96	0.07	-	98,98,98,98	0
8	MN	J	1009	1/1	0.76	0.12	-	113,113,113,113	0
7	CL	A	2001	1/1	0.99	0.13	-	57,57,57,57	0
8	MN	I	1007	1/1	0.94	0.06	-	131,131,131,131	0
8	MN	J	1012	1/1	0.57	0.09	-	128,128,128,128	0
8	MN	J	1015	1/1	0.77	0.13	-	105,105,105,105	0
7	CL	E	2002	1/1	0.97	0.12	-	56,56,56,56	0
8	MN	I	1008	1/1	0.90	0.22	-	81,81,81,81	0
8	MN	J	1002	1/1	0.96	0.22	-	71,71,71,71	0
8	MN	I	1017	1/1	0.99	0.05	-	96,96,96,96	0

### 6.5 Other polymers [i](#)

There are no such residues in this entry.