



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 02:17 PM GMT

PDB ID : 3WKJ  
Title : The nucleosome containing human TSH2B  
Authors : Urahama, T.; Horikoshi, N.; Osakabe, A.; Tachiwana, H.; Kurumizaka, H.  
Deposited on : 2013-10-22  
Resolution : 2.80 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  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
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) : trunk26865

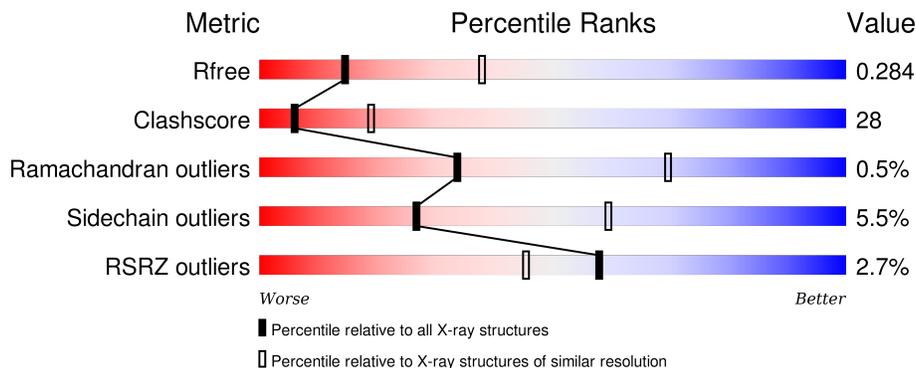
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.80 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))
$R_{free}$	91344	2393 (2.80-2.80)
Clashscore	102246	2827 (2.80-2.80)
Ramachandran outliers	100387	2782 (2.80-2.80)
Sidechain outliers	100360	2784 (2.80-2.80)
RSRZ outliers	91569	2404 (2.80-2.80)

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	139	 2% 49% 19% 30%
1	E	139	 2% 50% 19% 29%
2	B	106	 54% 19% 26%
2	F	106	 2% 59% 20% 21%
3	C	133	 2% 56% 20% 5% 20%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	G	133	
4	D	130	
4	H	130	
5	I	146	
5	J	146	

## 2 Entry composition [i](#)

There are 8 unique types of molecules in this entry. The entry contains 12026 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 H3.1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	97	801	505	155	137	4	0	0	0
1	E	98	806	508	156	138	4	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-3	GLY	-	EXPRESSION TAG	UNP P68431
A	-2	SER	-	EXPRESSION TAG	UNP P68431
A	-1	HIS	-	EXPRESSION TAG	UNP P68431
E	-3	GLY	-	EXPRESSION TAG	UNP P68431
E	-2	SER	-	EXPRESSION TAG	UNP P68431
E	-1	HIS	-	EXPRESSION TAG	UNP P68431

- Molecule 2 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	78	619	391	120	107	1	0	0	0
2	F	84	673	424	133	115	1	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-3	GLY	-	EXPRESSION TAG	UNP P62805
B	-2	SER	-	EXPRESSION TAG	UNP P62805
B	-1	HIS	-	EXPRESSION TAG	UNP P62805
F	-3	GLY	-	EXPRESSION TAG	UNP P62805
F	-2	SER	-	EXPRESSION TAG	UNP P62805
F	-1	HIS	-	EXPRESSION TAG	UNP P62805

- Molecule 3 is a protein called Histone H2A type 1-B/E.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
3	C	107	Total	C	N	O	0	0	0
			824	520	161	143			
3	G	104	Total	C	N	O	0	0	0
			805	508	157	140			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-3	GLY	-	EXPRESSION TAG	UNP P04908
C	-2	SER	-	EXPRESSION TAG	UNP P04908
C	-1	HIS	-	EXPRESSION TAG	UNP P04908
G	-3	GLY	-	EXPRESSION TAG	UNP P04908
G	-2	SER	-	EXPRESSION TAG	UNP P04908
G	-1	HIS	-	EXPRESSION TAG	UNP P04908

- Molecule 4 is a protein called Histone H2B type 1-A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	D	93	Total	C	N	O	S	0	0	0
			728	458	128	140	2			
4	H	92	Total	C	N	O	S	0	0	0
			721	454	127	138	2			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	-3	GLY	-	EXPRESSION TAG	UNP Q96A08
D	-2	SER	-	EXPRESSION TAG	UNP Q96A08
D	-1	HIS	-	EXPRESSION TAG	UNP Q96A08
H	-3	GLY	-	EXPRESSION TAG	UNP Q96A08
H	-2	SER	-	EXPRESSION TAG	UNP Q96A08
H	-1	HIS	-	EXPRESSION TAG	UNP Q96A08

- Molecule 5 is a DNA chain called DNA (145-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
5	I	145	Total	C	N	O	P	0	0	0
			2970	1421	538	867	144			
5	J	145	Total	C	N	O	P	0	0	0
			2969	1421	535	869	144			

- Molecule 6 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

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

- Molecule 7 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	J	2	Total 2	Mn 2	0	0
7	I	3	Total 3	Mn 3	0	0

- Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	13	Total 13	O 13	0	0
8	B	6	Total 6	O 6	0	0
8	C	11	Total 11	O 11	0	0
8	D	7	Total 7	O 7	0	0
8	E	22	Total 22	O 22	0	0
8	F	19	Total 19	O 19	0	0
8	G	11	Total 11	O 11	0	0
8	H	4	Total 4	O 4	0	0
8	I	1	Total 1	O 1	0	0
8	J	7	Total 7	O 7	0	0

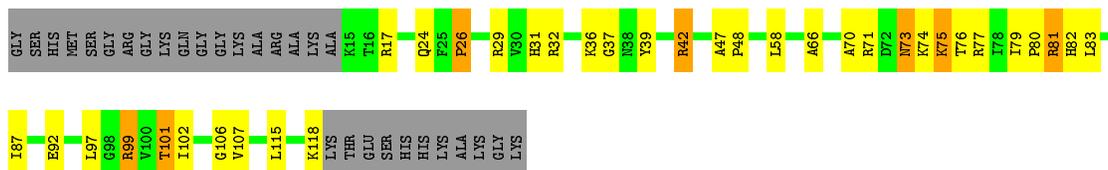


G102

Molecule 3: Histone H2A type 1-B/E



Molecule 3: Histone H2A type 1-B/E



Molecule 4: Histone H2B type 1-A

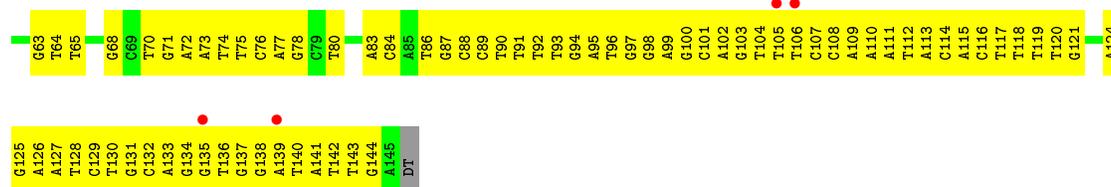


Molecule 4: Histone H2B type 1-A



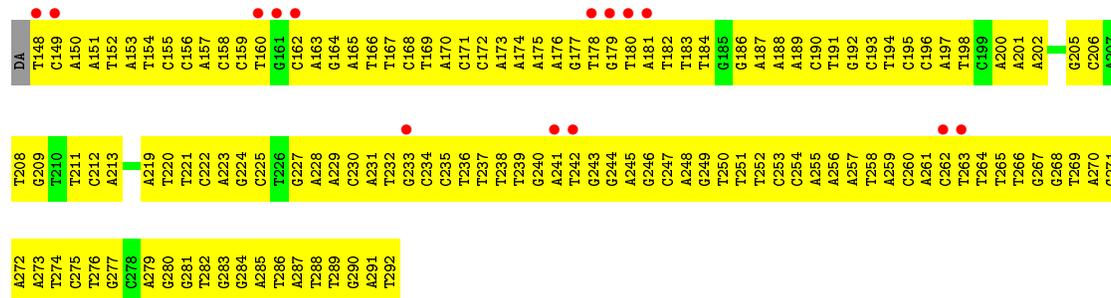
Molecule 5: DNA (145-MER)





● Molecule 5: DNA (145-MER)

Chain J: 10% 10% 90%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	107.04Å 109.93Å 182.96Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.80 47.24 – 2.80	Depositor EDS
% Data completeness (in resolution range)	(Not available) (50.00-2.80) 99.0 (47.24-2.80)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.08	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.33 (at 2.81Å)	Xtriage
Refinement program	CNS 1.21	Depositor
R, $R_{free}$	0.235 , 0.291 0.234 , 0.284	Depositor DCC
$R_{free}$ test set	2722 reflections (5.08%)	DCC
Wilson B-factor (Å <sup>2</sup> )	57.4	Xtriage
Anisotropy	0.471	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 55.0	EDS
Estimated twinning fraction	0.018 for k,h,l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 53611 reflections	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	12026	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	78.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.30% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.46	0/813	0.64	0/1090
1	E	0.61	1/818 (0.1%)	0.84	2/1097 (0.2%)
2	B	0.44	0/626	0.67	0/837
2	F	0.54	0/680	0.70	0/908
3	C	0.44	0/834	0.62	0/1125
3	G	0.38	0/815	0.60	0/1100
4	D	0.49	0/739	0.71	1/994 (0.1%)
4	H	0.45	0/732	0.61	0/984
5	I	0.38	0/3332	0.79	0/5141
5	J	0.37	0/3330	0.77	0/5138
All	All	0.43	1/12719 (0.0%)	0.74	3/18414 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	77	ASP	CB-CG	7.46	1.67	1.51

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	E	77	ASP	CB-CG-OD1	13.01	130.01	118.30
4	D	105	GLY	N-CA-C	-6.26	97.44	113.10
1	E	77	ASP	CB-CG-OD2	-5.30	113.53	118.30

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	801	0	839	39	0
1	E	806	0	844	31	0
2	B	619	0	659	24	0
2	F	673	0	722	21	0
3	C	824	0	884	40	0
3	G	805	0	861	40	0
4	D	728	0	751	44	0
4	H	721	0	744	37	0
5	I	2970	0	1640	207	0
5	J	2969	0	1641	219	0
6	A	1	0	0	0	0
6	C	1	0	0	0	0
6	E	1	0	0	0	0
6	G	1	0	0	0	0
7	I	3	0	0	0	0
7	J	2	0	0	0	0
8	A	13	0	0	1	0
8	B	6	0	0	0	0
8	C	11	0	0	0	0
8	D	7	0	0	1	0
8	E	22	0	0	3	0
8	F	19	0	0	2	0
8	G	11	0	0	1	0
8	H	4	0	0	0	0
8	I	1	0	0	0	0
8	J	7	0	0	0	0
All	All	12026	0	9585	612	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 28.

The worst 5 of 612 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:51:DA:H2"	5:I:52:DT:H5"	1.22	1.11

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
5:I:36:DT:H2''	5:I:37:DT:H5''	1.29	1.10
5:I:101:DC:H2''	5:I:102:DA:H5'	1.36	1.07
5:J:152:DT:H2''	5:J:153:DA:H5''	1.36	1.07
5:I:50:DC:H2''	5:I:51:DA:H5'	1.30	1.06

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	95/139 (68%)	93 (98%)	2 (2%)	0	100	100
1	E	96/139 (69%)	92 (96%)	3 (3%)	1 (1%)	19	52
2	B	76/106 (72%)	74 (97%)	2 (3%)	0	100	100
2	F	82/106 (77%)	79 (96%)	3 (4%)	0	100	100
3	C	105/133 (79%)	101 (96%)	3 (3%)	1 (1%)	19	52
3	G	102/133 (77%)	97 (95%)	4 (4%)	1 (1%)	19	52
4	D	91/130 (70%)	82 (90%)	8 (9%)	1 (1%)	17	50
4	H	90/130 (69%)	86 (96%)	4 (4%)	0	100	100
All	All	737/1016 (72%)	704 (96%)	29 (4%)	4 (0%)	34	69

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	64	LYS
4	D	86	LYS
3	C	114	VAL
3	G	26	PRO

### 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	85/113 (75%)	83 (98%)	2 (2%)	57	87
1	E	85/113 (75%)	80 (94%)	5 (6%)	24	57
2	B	63/81 (78%)	61 (97%)	2 (3%)	46	80
2	F	69/81 (85%)	67 (97%)	2 (3%)	50	83
3	C	84/102 (82%)	77 (92%)	7 (8%)	14	38
3	G	83/102 (81%)	77 (93%)	6 (7%)	18	45
4	D	82/113 (73%)	77 (94%)	5 (6%)	23	55
4	H	81/113 (72%)	75 (93%)	6 (7%)	17	43
All	All	632/818 (77%)	597 (94%)	35 (6%)	27	59

5 of 35 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	D	107	LEU
1	E	117	VAL
4	H	68	THR
1	E	49	ARG
1	E	64	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
4	D	50	HIS
4	D	96	GLN
2	F	93	GLN
3	C	84	GLN
1	E	76	GLN

### 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 9 ligands modelled in this entry, 9 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	97/139 (69%)	0.12	2 (2%) 67 56	28, 47, 71, 98	0
1	E	98/139 (70%)	-0.00	1 (1%) 84 77	26, 37, 61, 80	0
2	B	78/106 (73%)	0.01	0 100 100	32, 45, 59, 76	0
2	F	84/106 (79%)	0.06	2 (2%) 62 50	25, 36, 53, 79	0
3	C	107/133 (80%)	0.04	2 (1%) 70 59	26, 43, 69, 100	0
3	G	104/133 (78%)	-0.05	0 100 100	33, 51, 75, 86	0
4	D	93/130 (71%)	0.12	0 100 100	27, 44, 69, 95	0
4	H	92/130 (70%)	0.20	0 100 100	29, 51, 76, 89	0
5	I	145/146 (99%)	0.70	7 (4%) 34 23	52, 111, 159, 177	0
5	J	145/146 (99%)	0.75	14 (9%) 10 5	54, 111, 155, 163	0
All	All	1043/1308 (79%)	0.24	28 (2%) 58 45	25, 51, 139, 177	0

The worst 5 of 28 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	J	162	DC	5.0
5	J	241	DA	4.9
2	F	102	GLY	4.7
5	I	55	DA	4.5
5	J	148	DT	4.3

### 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(Å <sup>2</sup> )	Q<0.9
7	MN	I	403	1/1	0.88	0.16	-1.33	95,95,95,95	0
6	CL	G	201	1/1	0.95	0.11	-2.08	50,50,50,50	0
6	CL	C	201	1/1	0.98	0.09	-2.40	48,48,48,48	0
7	MN	I	401	1/1	0.70	0.12	-	115,115,115,115	0
6	CL	A	201	1/1	0.92	0.07	-	60,60,60,60	0
7	MN	J	302	1/1	0.95	0.04	-	84,84,84,84	0
6	CL	E	201	1/1	0.90	0.11	-	58,58,58,58	0
7	MN	I	402	1/1	0.88	0.09	-	121,121,121,121	0
7	MN	J	301	1/1	0.77	0.14	-	99,99,99,99	0

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

There are no such residues in this entry.