wwPDB X-ray Structure Validation Summary Report (i)

Sep 25, 2023 – 04:28 AM EDT

PDB ID	:	5WH1
Title	:	Apo form of the C-terminal region of human Transcription Factor IIB
Authors	:	Bratkowski, M.A.; Liu, X.
Deposited on	:	2017-07-14
Resolution	:	3.39 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:

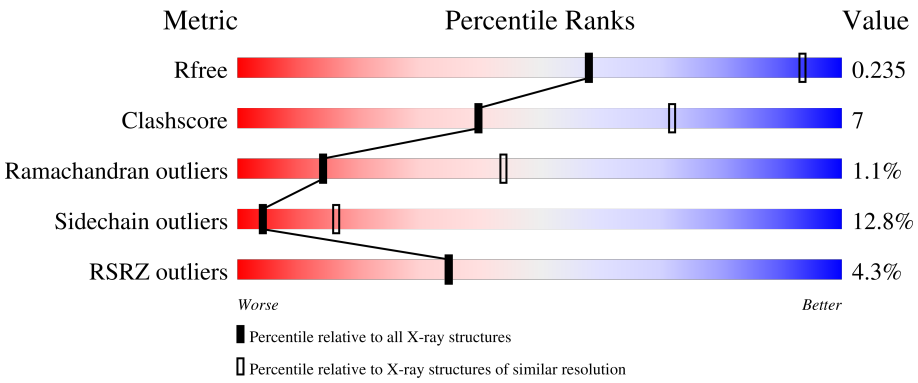
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 3.39 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	1026 (3.48-3.32)
Clashscore	141614	1055 (3.48-3.32)
Ramachandran outliers	138981	1038 (3.48-3.32)
Sidechain outliers	138945	1038 (3.48-3.32)
RSRZ outliers	127900	2173 (3.50-3.30)

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	211	5%, 71%, 23%, • •
1	B	211	10%, 75%, 20%, •
1	C	211	76%, 21%, •
1	D	211	%, 75%, 22%, •

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 6640 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.

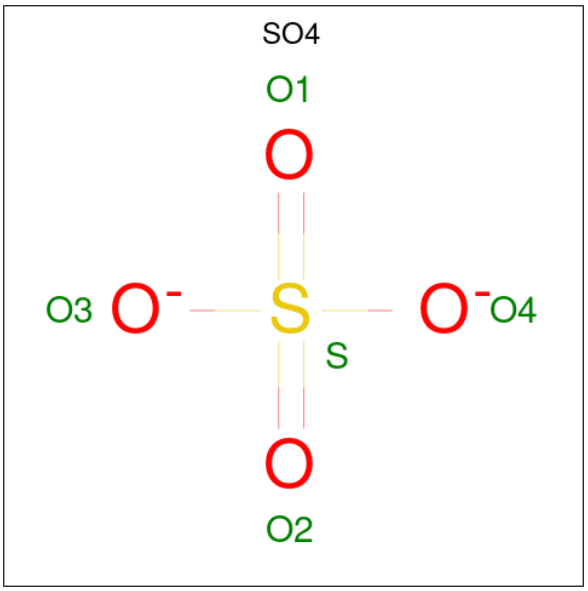
- Molecule 1 is a protein called Transcription initiation factor IIB.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	206	1605	1009	287	297	12	0	0	0
1	B	210	1635	1028	291	303	13	0	0	0
1	C	210	1635	1028	291	303	13	0	0	0
1	D	210	1635	1028	291	303	13	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	106	SER	-	expression tag	UNP Q00403
B	106	SER	-	expression tag	UNP Q00403
C	106	SER	-	expression tag	UNP Q00403
D	106	SER	-	expression tag	UNP Q00403

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	O	S			
2	A	1	5	4	1		0	0
			Total	O	S			
2	A	1	5	4	1		0	0
			Total	O	S			
2	A	1	5	4	1		0	0
			Total	O	S			
2	A	1	5	4	1		0	0
			Total	O	S			
2	A	1	5	4	1		0	0
			Total	O	S			
2	B	1	5	4	1		0	0
			Total	O	S			
2	B	1	5	4	1		0	0
			Total	O	S			
2	B	1	5	4	1		0	0
			Total	O	S			
2	B	1	5	4	1		0	0
			Total	O	S			
2	B	1	5	4	1		0	0
			Total	O	S			
2	C	1	5	4	1		0	0
			Total	O	S			
2	C	1	5	4	1		0	0
			Total	O	S			
2	C	1	5	4	1		0	0
			Total	O	S			
2	C	1	5	4	1		0	0

Continued on next page...

Continued from previous page...

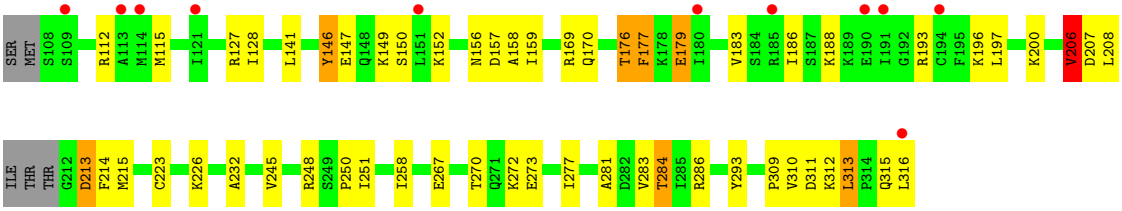
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	O	S		
2	C	1	5	4	1	0	0
			Total	O	S		
2	C	1	5	4	1	0	0
			Total	O	S		
2	C	1	5	4	1	0	0
			Total	O	S		
2	C	1	5	4	1	0	0
			Total	O	S		
2	D	1	5	4	1	0	0
			Total	O	S		
2	D	1	5	4	1	0	0
			Total	O	S		
2	D	1	5	4	1	0	0
			Total	O	S		
2	D	1	5	4	1	0	0
			Total	O	S		
2	D	1	5	4	1	0	0
			Total	O	S		
2	D	1	5	4	1	0	0
			Total	O	S		
2	D	1	5	4	1	0	0
			Total	O	S		
2	D	1	5	4	1	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.

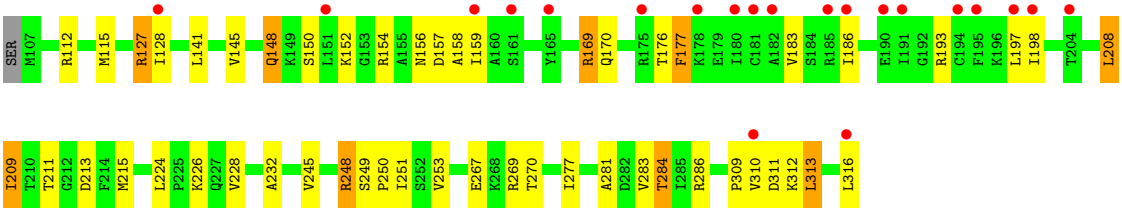
• Molecule 1: Transcription initiation factor IIB

Chain A:



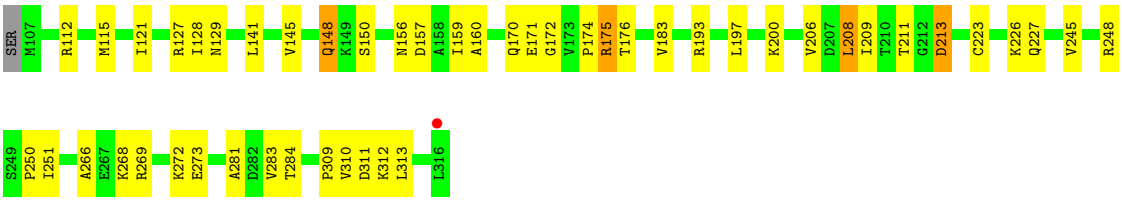
• Molecule 1: Transcription initiation factor IIB

Chain B:



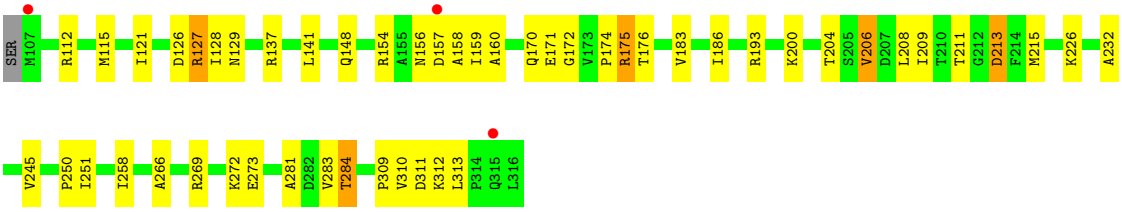
• Molecule 1: Transcription initiation factor IIB

Chain C:



• Molecule 1: Transcription initiation factor IIB

Chain D:


SER
M107
R112
M115
I121
D126
R127
I128
N129
R137
L141
Q148
R154
A155
N156
D157
A158
I159
A160
Q170
E171
G172
V173
P174
R175
T176
V183
I186
R193
K200
T204
S205
V206
D207
L208
I209
T210
T211
G212
D213
F214
M215
K226
A232
V245
P250
I251
I258
A266
R269
K272
E273
A281
D282
V283
T284
P309
V310
D311
K312
L313
P314
Q315
L316

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	108.17Å 126.46Å 158.43Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.73 – 3.39 49.73 – 3.39	Depositor EDS
% Data completeness (in resolution range)	99.6 (49.73-3.39) 99.6 (49.73-3.39)	Depositor EDS
R_{merge}	0.24	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.56 (at 3.40Å)	Xtriage
Refinement program	BUSTER 2.10.3	Depositor
R, R_{free}	0.194 , 0.214 0.219 , 0.235	Depositor DCC
R_{free} test set	1511 reflections (4.92%)	wwPDB-VP
Wilson B-factor (Å^2)	92.4	Xtriage
Anisotropy	0.637	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 80.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.46$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	6640	wwPDB-VP
Average B, all atoms (Å^2)	104.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.12% 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: SO4

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	0.50	0/1628	0.73	0/2191
1	B	0.50	0/1659	0.75	0/2235
1	C	0.51	0/1659	0.73	0/2235
1	D	0.50	0/1659	0.73	0/2235
All	All	0.50	0/6605	0.73	0/8896

There are no bond length outliers.

There are no bond angle outliers.

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	1605	0	1654	29	0
1	B	1635	0	1689	32	0
1	C	1635	0	1689	22	0
1	D	1635	0	1689	28	0
2	A	25	0	0	0	0
2	B	25	0	0	0	0
2	C	40	0	0	0	0
2	D	40	0	0	0	0
All	All	6640	0	6721	98	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 7.

The worst 5 of 98 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:B:277:ILE:HD11	1:C:206:VAL:HG12	1.71	0.72
1:D:172:GLY:CA	1:D:251:ILE:HD13	2.22	0.70
1:B:245:VAL:HG22	1:B:248:ARG:HG3	1.73	0.69
1:C:281:ALA:O	1:C:284:THR:HG22	1.93	0.68
1:A:208:LEU:C	1:A:208:LEU:HD23	2.14	0.67

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	202/211 (96%)	192 (95%)	7 (4%)	3 (2%)	10	36
1	B	208/211 (99%)	197 (95%)	9 (4%)	2 (1%)	15	46
1	C	208/211 (99%)	199 (96%)	7 (3%)	2 (1%)	15	46
1	D	208/211 (99%)	200 (96%)	6 (3%)	2 (1%)	15	46
All	All	826/844 (98%)	788 (95%)	29 (4%)	9 (1%)	14	44

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	206	VAL
1	B	148	GLN
1	B	211	THR
1	C	148	GLN
1	C	211	THR

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	175/180 (97%)	151 (86%)	24 (14%)	3	14
1	B	179/180 (99%)	158 (88%)	21 (12%)	5	20
1	C	179/180 (99%)	156 (87%)	23 (13%)	4	16
1	D	179/180 (99%)	156 (87%)	23 (13%)	4	16
All	All	712/720 (99%)	621 (87%)	91 (13%)	4	16

5 of 91 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	209	ILE
1	D	127	ARG
1	C	226	LYS
1	C	310	VAL
1	D	170	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	221	ASN
1	B	139	ASN
1	D	139	ASN
1	B	227	GLN
1	A	148	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

26 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	$\#\|Z\| > 2$
2	SO4	D	407	-	4,4,4	0.23	0	6,6,6	0.15	0
2	SO4	C	408	-	4,4,4	0.15	0	6,6,6	0.18	0
2	SO4	D	402	-	4,4,4	0.08	0	6,6,6	0.24	0
2	SO4	D	404	-	4,4,4	0.18	0	6,6,6	0.07	0
2	SO4	C	407	-	4,4,4	0.13	0	6,6,6	0.24	0
2	SO4	B	401	-	4,4,4	0.20	0	6,6,6	0.20	0
2	SO4	C	405	-	4,4,4	0.15	0	6,6,6	0.25	0
2	SO4	B	403	-	4,4,4	0.12	0	6,6,6	0.19	0
2	SO4	A	401	-	4,4,4	0.22	0	6,6,6	0.19	0
2	SO4	C	406	-	4,4,4	0.22	0	6,6,6	0.10	0
2	SO4	A	402	-	4,4,4	0.16	0	6,6,6	0.21	0
2	SO4	B	405	-	4,4,4	0.23	0	6,6,6	0.21	0
2	SO4	A	404	-	4,4,4	0.22	0	6,6,6	0.19	0
2	SO4	C	403	-	4,4,4	0.21	0	6,6,6	0.26	0
2	SO4	D	401	-	4,4,4	0.37	0	6,6,6	0.24	0
2	SO4	D	403	-	4,4,4	0.19	0	6,6,6	0.10	0
2	SO4	D	406	-	4,4,4	0.26	0	6,6,6	0.18	0
2	SO4	A	405	-	4,4,4	0.10	0	6,6,6	0.37	0
2	SO4	C	402	-	4,4,4	0.16	0	6,6,6	0.21	0
2	SO4	D	405	-	4,4,4	0.20	0	6,6,6	0.24	0
2	SO4	C	404	-	4,4,4	0.12	0	6,6,6	0.18	0
2	SO4	B	402	-	4,4,4	0.27	0	6,6,6	0.36	0
2	SO4	B	404	-	4,4,4	0.10	0	6,6,6	0.28	0
2	SO4	D	408	-	4,4,4	0.20	0	6,6,6	0.24	0
2	SO4	A	403	-	4,4,4	0.10	0	6,6,6	0.53	0
2	SO4	C	401	-	4,4,4	0.08	0	6,6,6	0.26	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.

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, 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	206/211 (97%)	0.49	11 (5%) 26 27	68, 107, 163, 200	0
1	B	210/211 (99%)	0.62	21 (10%) 7 8	68, 106, 153, 169	0
1	C	210/211 (99%)	0.17	1 (0%) 91 90	69, 91, 122, 144	0
1	D	210/211 (99%)	0.24	3 (1%) 75 74	71, 91, 125, 145	0
All	All	836/844 (99%)	0.38	36 (4%) 35 35	68, 96, 150, 200	0

The worst 5 of 36 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	151	LEU	4.1
1	B	180	ILE	4.0
1	A	316	LEU	3.8
1	B	190	GLU	3.8
1	B	204	THR	3.7

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
2	SO4	C	407	5/5	0.53	0.31	180,182,183,185	0
2	SO4	A	404	5/5	0.69	0.29	162,164,164,164	0
2	SO4	D	407	5/5	0.70	0.32	155,157,159,159	0
2	SO4	D	406	5/5	0.72	0.39	183,183,184,185	0
2	SO4	B	403	5/5	0.72	0.39	154,156,157,157	0
2	SO4	B	405	5/5	0.75	0.29	172,173,173,174	0
2	SO4	C	406	5/5	0.76	0.33	151,153,154,155	0
2	SO4	A	401	5/5	0.78	0.26	176,176,178,179	0
2	SO4	D	404	5/5	0.78	0.25	180,182,183,183	0
2	SO4	D	403	5/5	0.79	0.34	158,160,161,161	0
2	SO4	C	404	5/5	0.83	0.26	180,180,181,182	0
2	SO4	C	408	5/5	0.83	0.34	194,194,195,195	0
2	SO4	C	405	5/5	0.84	0.21	153,154,156,156	0
2	SO4	D	405	5/5	0.84	0.15	163,164,165,167	0
2	SO4	D	408	5/5	0.84	0.33	181,182,183,185	0
2	SO4	A	405	5/5	0.85	0.16	150,153,153,154	0
2	SO4	B	401	5/5	0.86	0.19	157,157,159,159	0
2	SO4	B	404	5/5	0.86	0.30	150,151,153,153	0
2	SO4	A	402	5/5	0.89	0.19	158,160,161,162	0
2	SO4	C	403	5/5	0.90	0.20	156,157,159,160	0
2	SO4	B	402	5/5	0.93	0.17	104,106,107,108	0
2	SO4	D	401	5/5	0.93	0.16	103,104,106,106	0
2	SO4	C	402	5/5	0.94	0.13	119,122,123,126	0
2	SO4	A	403	5/5	0.94	0.17	100,100,100,101	0
2	SO4	D	402	5/5	0.94	0.12	132,133,134,135	0
2	SO4	C	401	5/5	0.95	0.20	107,107,107,108	0

6.5 Other polymers (i)

There are no such residues in this entry.