Full wwPDB X-ray Structure Validation Report (i)

May 21, 2020 – 12:37 am BST

PDB ID	:	3WNE
Title	:	Cyclic hexapeptide PKIDNG in complex with HIV-1 integrase
Authors	:	Wielens, J.; Chalmers, D.K.; Parker, M.W.
Deposited on	:	2013-12-09
Resolution	:	1.70 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 (i)) were used in the production of this report:

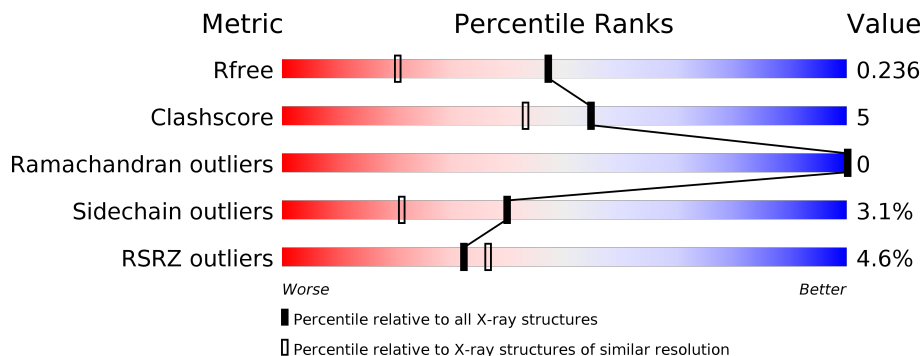|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry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

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.70 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	Percentile Ranks	Value
Rfree		0.236
Clashscore		5
Ramachandran outliers		0
Sidechain outliers		3.1%
RSRZ outliers		4.6%

Worse   Better

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	4298 (1.70-1.70)
Clashscore	141614	4695 (1.70-1.70)
Ramachandran outliers	138981	4610 (1.70-1.70)
Sidechain outliers	138945	4610 (1.70-1.70)
RSRZ outliers	127900	4222 (1.70-1.70)

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	157	2% / 82% / 6% · 12%
1	B	157	2% / 78% / 6% / 16%
2	C	6	50% / 17% / 67% / 17%
2	D	6	67% / 33% / 33% / 33%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 2370 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 Gag-Pol polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	138	Total	C	N	O	S	0	4	0
			1086	686	193	203	4			
1	B	132	Total	C	N	O	S	0	3	0
			1043	663	184	192	4			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	56	SER	CYS	ENGINEERED MUTATION	UNP P12497
A	123	GLY	SER	ENGINEERED MUTATION	UNP P12497
A	124	ALA	THR	ENGINEERED MUTATION	UNP P12497
A	127	ARG	LYS	ENGINEERED MUTATION	UNP P12497
A	131	ASP	TRP	ENGINEERED MUTATION	UNP P12497
A	139	ASP	PHE	ENGINEERED MUTATION	UNP P12497
A	185	HIS	PHE	ENGINEERED MUTATION	UNP P12497
B	56	SER	CYS	ENGINEERED MUTATION	UNP P12497
B	123	GLY	SER	ENGINEERED MUTATION	UNP P12497
B	124	ALA	THR	ENGINEERED MUTATION	UNP P12497
B	127	ARG	LYS	ENGINEERED MUTATION	UNP P12497
B	131	ASP	TRP	ENGINEERED MUTATION	UNP P12497
B	139	ASP	PHE	ENGINEERED MUTATION	UNP P12497
B	185	HIS	PHE	ENGINEERED MUTATION	UNP P12497

- Molecule 2 is a protein called LEDGF peptide.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	C	6	Total	C	N	O	0	0	0
			44	27	8	9			
2	D	6	Total	C	N	O	0	0	0
			44	27	8	9			

- Molecule 3 is CADMIUM ION (three-letter code: CD) (formula: Cd).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	2	Total 2	Cd 2	0	0
3	A	2	Total 2	Cd 2	0	0

- Molecule 4 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total 1	Cl 1	0	0
4	A	1	Total 1	Cl 1	0	0

- Molecule 5 is SULFATE ION (three-letter code: SO4) (formula: O_4S).
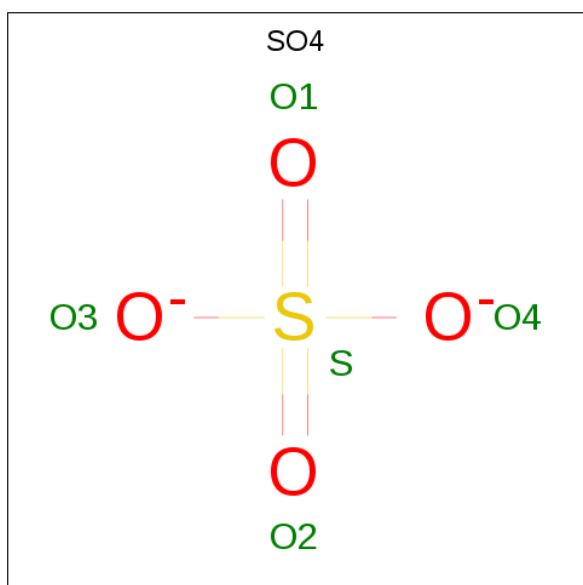
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total 5	O 4	S 1	0	0
5	A	1	Total 5	O 4	S 1	0	0
5	A	1	Total 5	O 4	S 1	0	0
5	B	1	Total 5	O 4	S 1	0	0
5	B	1	Total 5	O 4	S 1	0	0
5	B	1	Total 5	O 4	S 1	0	0

- Molecule 6 is water.

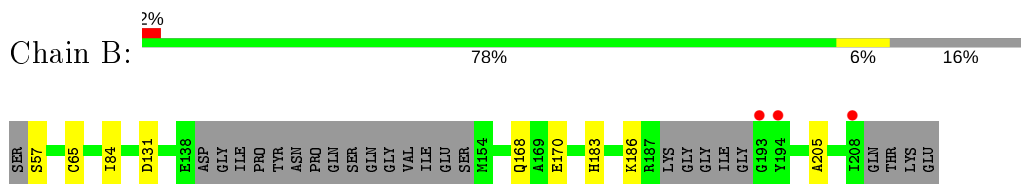| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	55	Total 55	O 55	0	0
6	B	62	Total 62	O 62	0	0

3 Residue-property plots ⓘ

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.
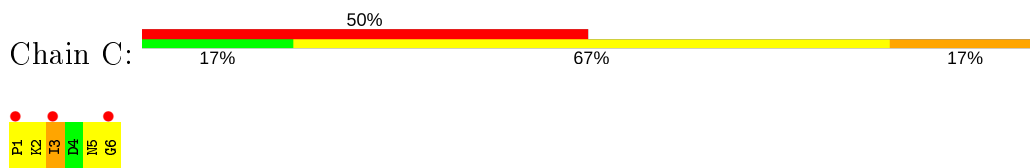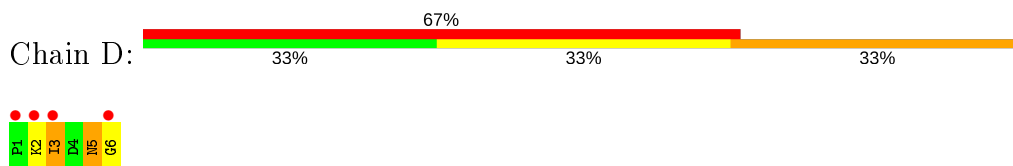
- Molecule 1: Gag-Pol polyprotein
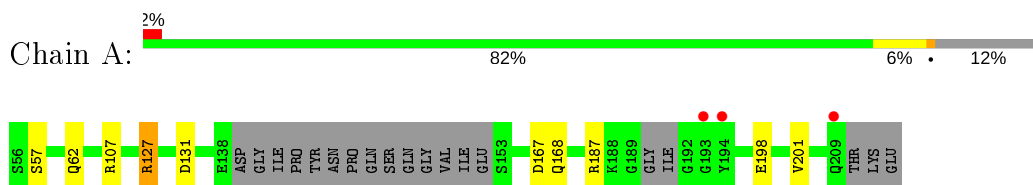
Chain A:

- Molecule 1: Gag-Pol polyprotein

Chain B:

- Molecule 2: LEDGF peptide

Chain C:

- Molecule 2: LEDGF peptide

Chain D:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32	Depositor
Cell constants a, b, c, α, β, γ	49.68Å 49.68Å 102.93Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	39.70 – 1.70 39.70 – 1.70	Depositor EDS
% Data completeness (in resolution range)	100.0 (39.70-1.70) 99.0 (39.70-1.70)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	2.37 (at 1.70Å)	Xtriage
Refinement program	REFMAC 5.5.0072	Depositor
R, R_{free}	0.195 , 0.228 0.199 , 0.236	Depositor DCC
R_{free} test set	1547 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å²)	25.0	Xtriage
Anisotropy	0.011	Xtriage
Bulk solvent k_{sol}(e/Å³), B_{sol}(Å²)	0.37 , 31.2	EDS
L-test for twinning[2]	$<|L|> = 0.50$, $<L^2> = 0.34$	Xtriage
Estimated twinning fraction	0.025 for -h,-k,l 0.488 for h,-h-k,-l 0.026 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	2370	wwPDB-VP
Average B, all atoms (Å²)	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.98% 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: CL, SO4, CD

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	1.06	0/1104	0.92	2/1493 (0.1%)
1	B	1.07	2/1061 (0.2%)	0.92	2/1436 (0.1%)
2	C	1.17	0/44	1.33	0/57
2	D	1.22	0/44	1.56	0/57
All	All	1.07	2/2253 (0.1%)	0.95	4/3043 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	65	CYS	CB-SG	6.27	1.93	1.82
1	B	131	ASP	CB-CG	5.67	1.63	1.51

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	131	ASP	CB-CG-OD1	-7.45	111.60	118.30
1	A	127	ARG	NE-CZ-NH1	-6.55	117.03	120.30
1	A	131	ASP	CB-CG-OD1	-6.39	112.55	118.30
1	B	131	ASP	OD1-CG-OD2	5.57	133.88	123.30

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	1086	0	1073	5	0
1	B	1043	0	1040	5	0
2	C	44	0	44	9	0
2	D	44	0	44	9	0
3	A	2	0	0	0	0
3	B	2	0	0	0	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
5	A	15	0	0	0	0
5	B	15	0	0	0	0
6	A	55	0	0	0	0
6	B	62	0	0	1	0
All	All	2370	0	2201	23	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 5.

All (23) 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:D:5:ASN:CG	2:D:6:GLY:H	1.67	0.98
2:C:1:PRO:HD2	2:C:5:ASN:ND2	1.85	0.91
2:C:1:PRO:HD2	2:C:5:ASN:HD21	1.37	0.89
2:C:5:ASN:CG	2:C:6:GLY:N	2.29	0.84
2:D:5:ASN:CG	2:D:6:GLY:N	2.30	0.83
2:D:5:ASN:ND2	2:D:6:GLY:N	2.29	0.81
2:D:5:ASN:ND2	2:D:6:GLY:H	1.88	0.69
2:C:5:ASN:CG	2:C:6:GLY:H	1.96	0.67
1:B:170:GLU:HB2	2:D:5:ASN:OD1	1.95	0.65
1:A:187:ARG:NH2	1:A:198:GLU:OE2	2.36	0.58
2:C:5:ASN:O	2:C:6:GLY:C	2.43	0.56
1:B:57:SER:OG	6:B:444:HOH:O	2.19	0.53
1:B:84[A]:ILE:HD12	1:B:183:HIS:CD2	2.44	0.52
2:C:5:ASN:O	2:C:6:GLY:O	2.30	0.50
2:C:2:LYS:O	2:C:5:ASN:O	2.30	0.49
1:A:201:VAL:HG12	1:B:205:ALA:HB2	1.95	0.48
2:D:2:LYS:H	2:D:6:GLY:C	2.16	0.47
2:D:2:LYS:N	2:D:6:GLY:C	2.68	0.47
2:D:5:ASN:ND2	2:D:6:GLY:CA	2.80	0.44
1:B:168:GLN:O	2:D:3:ILE:HB	2.18	0.43
1:A:168:GLN:O	2:C:3:ILE:HB	2.18	0.43
1:A:167:ASP:O	2:C:2:LYS:HD2	2.20	0.41

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:107:ARG:HG3	1:A:107:ARG:HH11	1.86	0.40

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	136/157 (87%)	132 (97%)	4 (3%)	0	100	100
1	B	129/157 (82%)	128 (99%)	1 (1%)	0	100	100
2	C	4/6 (67%)	3 (75%)	1 (25%)	0	100	100
2	D	4/6 (67%)	3 (75%)	1 (25%)	0	100	100
All	All	273/326 (84%)	266 (97%)	7 (3%)	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	112/126 (89%)	109 (97%)	3 (3%)	44	26
1	B	108/126 (86%)	107 (99%)	1 (1%)	78	70
2	C	5/5 (100%)	4 (80%)	1 (20%)	1	0
2	D	5/5 (100%)	3 (60%)	2 (40%)	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	230/262 (88%)	223 (97%)	7 (3%)	40　22

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	57	SER
1	A	62	GLN
1	A	127	ARG
1	B	186	LYS
2	C	3	ILE
2	D	3	ILE
2	D	5	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	155	ASN
1	B	164	GLN

5.3.3　RNA ⓘ

There are no RNA molecules in this entry.

5.4　Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5　Carbohydrates ⓘ

There are no carbohydrates in this entry.

5.6　Ligand geometry ⓘ

Of 12 ligands modelled in this entry, 6 are monoatomic - leaving 6 for Mogul analysis.

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	$\#|Z| > 2$
5	SO4	B	305	-	4,4,4	0.29	0	6,6,6	0.62	0
5	SO4	B	304	-	4,4,4	0.13	0	6,6,6	0.57	0
5	SO4	A	1006	-	4,4,4	0.38	0	6,6,6	0.40	0
5	SO4	B	306	-	4,4,4	0.25	0	6,6,6	0.28	0
5	SO4	A	1005	-	4,4,4	0.24	0	6,6,6	0.40	0
5	SO4	A	1004	-	4,4,4	0.35	0	6,6,6	0.63	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	138/157 (87%)	-0.13	3 (2%)	62	66	17, 26, 54, 64	0
1	B	132/157 (84%)	-0.22	3 (2%)	60	65	17, 26, 46, 53	0
2	C	6/6 (100%)	2.33	3 (50%)	0	0	42, 60, 72, 72	0
2	D	6/6 (100%)	1.95	4 (66%)	0	0	45, 61, 69, 71	0
All	All	282/326 (86%)	-0.07	13 (4%)	32	36	17, 27, 54, 72	0

All (13) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	1	PRO	6.4
1	A	193	GLY	5.6
1	B	193	GLY	4.5
1	A	194	TYR	3.5
2	C	3	ILE	3.3
2	C	6	GLY	3.1
2	D	3	ILE	3.0
1	B	208	ILE	2.9
1	B	194	TYR	2.9
2	D	1	PRO	2.9
2	D	6	GLY	2.2
1	A	209	GLN	2.1
2	D	2	LYS	2.1

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
5	SO4	B	305	5/5	0.90	0.13	45,46,48,49	5
5	SO4	A	1005	5/5	0.93	0.10	42,42,45,46	5
5	SO4	A	1004	5/5	0.94	0.10	29,43,44,46	0
5	SO4	B	304	5/5	0.95	0.09	30,45,46,48	0
5	SO4	A	1006	5/5	0.97	0.10	46,49,50,54	0
5	SO4	B	306	5/5	0.97	0.12	48,49,51,55	0
3	CD	B	301	1/1	0.98	0.08	23,23,23,23	0
4	CL	A	1003	1/1	0.98	0.05	23,23,23,23	0
3	CD	A	1002	1/1	0.99	0.07	23,23,23,23	0
3	CD	A	1001	1/1	0.99	0.09	23,23,23,23	0
3	CD	B	302	1/1	0.99	0.06	24,24,24,24	0
4	CL	B	303	1/1	0.99	0.05	24,24,24,24	0

6.5 Other polymers (i)

There are no such residues in this entry.