



wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 24, 2023 – 07:22 AM EDT

PDB ID : 5UN8
Title : Crystal Structure of human O-GlcNAcase in complex with glycopeptide p53
Authors : Li, B.; Jiang, J.
Deposited on : 2017-01-30
Resolution : 2.13 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 ⓘ 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 ⓘ](#)) 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

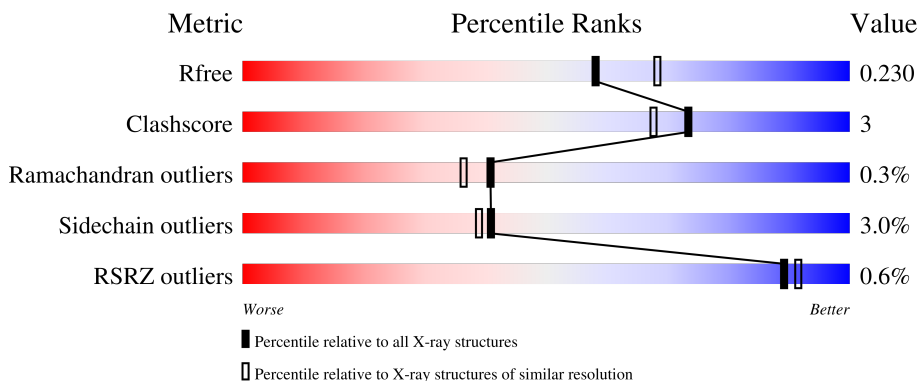
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.13 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	2523 (2.16-2.12)
Clashscore	141614	2653 (2.16-2.12)
Ramachandran outliers	138981	2618 (2.16-2.12)
Sidechain outliers	138945	2617 (2.16-2.12)
RSRZ outliers	127900	2485 (2.16-2.12)

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 $\leq 5\%$. 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	504	 76% 7% • 16%
1	B	504	 76% 8% • 15%
1	C	504	 77% 8% • 15%
1	D	504	 74% 11% • 15%
2	E	11	 55% 9% 36%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	11	 73% 27%
2	G	11	 82% 18%
2	H	11	 73% 27%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 15552 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 Protein O-GlcNAcase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	424	Total 3480	C 2254	N 575	O 628	S 23	0	0	0
1	B	426	Total 3495	C 2263	N 578	O 631	S 23	0	0	0
1	C	429	Total 3520	C 2279	N 581	O 636	S 24	0	0	0
1	D	429	Total 3520	C 2279	N 581	O 636	S 24	0	0	0

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	59	HIS	-	expression tag	UNP O60502
A	175	ASN	ASP	conflict	UNP O60502
A	543	GLY	-	linker	UNP O60502
A	544	GLY	-	linker	UNP O60502
A	545	GLY	-	linker	UNP O60502
A	546	GLY	-	linker	UNP O60502
A	547	SER	-	linker	UNP O60502
A	548	GLY	-	linker	UNP O60502
A	549	GLY	-	linker	UNP O60502
A	550	GLY	-	linker	UNP O60502
A	551	GLY	-	linker	UNP O60502
A	552	SER	-	linker	UNP O60502
B	59	HIS	-	expression tag	UNP O60502
B	175	ASN	ASP	conflict	UNP O60502
B	543	GLY	-	linker	UNP O60502
B	544	GLY	-	linker	UNP O60502
B	545	GLY	-	linker	UNP O60502
B	546	GLY	-	linker	UNP O60502
B	547	SER	-	linker	UNP O60502
B	548	GLY	-	linker	UNP O60502
B	549	GLY	-	linker	UNP O60502

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	550	GLY	-	linker	UNP O60502
B	551	GLY	-	linker	UNP O60502
B	552	SER	-	linker	UNP O60502
C	59	HIS	-	expression tag	UNP O60502
C	175	ASN	ASP	conflict	UNP O60502
C	543	GLY	-	linker	UNP O60502
C	544	GLY	-	linker	UNP O60502
C	545	GLY	-	linker	UNP O60502
C	546	GLY	-	linker	UNP O60502
C	547	SER	-	linker	UNP O60502
C	548	GLY	-	linker	UNP O60502
C	549	GLY	-	linker	UNP O60502
C	550	GLY	-	linker	UNP O60502
C	551	GLY	-	linker	UNP O60502
C	552	SER	-	linker	UNP O60502
D	59	HIS	-	expression tag	UNP O60502
D	175	ASN	ASP	conflict	UNP O60502
D	543	GLY	-	linker	UNP O60502
D	544	GLY	-	linker	UNP O60502
D	545	GLY	-	linker	UNP O60502
D	546	GLY	-	linker	UNP O60502
D	547	SER	-	linker	UNP O60502
D	548	GLY	-	linker	UNP O60502
D	549	GLY	-	linker	UNP O60502
D	550	GLY	-	linker	UNP O60502
D	551	GLY	-	linker	UNP O60502
D	552	SER	-	linker	UNP O60502

- Molecule 2 is a protein called P53 peptide.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	E	7	Total	C	N	O	0	0	0
			56	37	8	11			
2	F	8	Total	C	N	O	0	0	0
			63	42	9	12			
2	G	9	Total	C	N	O	0	0	0
			67	44	10	13			
2	H	8	Total	C	N	O	0	0	0
			63	42	9	12			

- Molecule 3 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: C₈H₁₅NO₆).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	E	1	Total	C	N	O	0	0
			14	8	1	5		
3	F	1	Total	C	N	O	0	0
			14	8	1	5		
3	G	1	Total	C	N	O	0	0
			14	8	1	5		
3	H	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	317	Total	O	0	0
			317	317		
4	B	313	Total	O	0	0
			313	313		
4	C	277	Total	O	0	0
			277	277		
4	D	289	Total	O	0	0
			289	289		
4	E	13	Total	O	0	0
			13	13		
4	F	8	Total	O	0	0
			8	8		
4	G	6	Total	O	0	0
			6	6		
4	H	9	Total	O	0	0
			9	9		

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	89.91Å 95.39Å 149.32Å 90.00° 96.91° 90.00°	Depositor
Resolution (Å)	148.24 – 2.13 49.41 – 2.13	Depositor EDS
% Data completeness (in resolution range)	96.8 (148.24-2.13) 96.8 (49.41-2.13)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.85 (at 2.14Å)	Xtrriage
Refinement program	REFMAC 5.8.0131	Depositor
R, R_{free}	0.184 , 0.229 0.185 , 0.230	Depositor DCC
R_{free} test set	6796 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å ²)	24.3	Xtrriage
Anisotropy	0.053	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 58.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	15552	wwPDB-VP
Average B, all atoms (Å ²)	30.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 39.09 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 3.3407e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹Intensities estimated from amplitudes.

²Theoretical 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: NAG

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	1.05	4/3574 (0.1%)	0.98	9/4838 (0.2%)
1	B	1.01	0/3589	0.92	7/4857 (0.1%)
1	C	1.00	1/3615 (0.0%)	0.91	7/4893 (0.1%)
1	D	0.98	1/3615 (0.0%)	0.96	10/4893 (0.2%)
2	E	1.23	0/59	0.97	0/83
2	F	1.15	0/67	1.02	0/95
2	G	1.25	0/71	0.84	0/100
2	H	1.22	0/67	0.92	0/95
All	All	1.01	6/14657 (0.0%)	0.94	33/19854 (0.2%)

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	663	CYS	CB-SG	12.03	2.02	1.82
1	A	663	CYS	CA-CB	6.01	1.67	1.53
1	A	663	CYS	N-CA	5.82	1.57	1.46
1	D	629	SER	CB-OG	-5.10	1.35	1.42
1	C	663	CYS	CA-CB	5.10	1.65	1.53

The worst 5 of 33 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	663	CYS	CA-CB-SG	19.58	149.24	114.00
1	D	691	ARG	NE-CZ-NH2	-13.86	113.37	120.30
1	B	83	ARG	NE-CZ-NH1	7.62	124.11	120.30
1	D	640	MET	CG-SD-CE	-7.46	88.26	100.20
1	B	83	ARG	NE-CZ-NH2	-6.98	116.81	120.30

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

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	3480	0	3401	19	1
1	B	3495	0	3419	30	1
1	C	3520	0	3446	21	1
1	D	3520	0	3446	31	0
2	E	56	0	47	1	0
2	F	63	0	54	0	0
2	G	67	0	57	0	0
2	H	63	0	54	0	0
3	E	14	0	13	0	0
3	F	14	0	13	0	0
3	G	14	0	13	0	0
3	H	14	0	13	0	0
4	A	317	0	0	4	0
4	B	313	0	0	8	0
4	C	277	0	0	7	0
4	D	289	0	0	5	1
4	E	13	0	0	0	0
4	F	8	0	0	0	0
4	G	6	0	0	0	0
4	H	9	0	0	0	0
All	All	15552	0	13976	94	2

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 3.

The worst 5 of 94 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:A:663:CYS:CB	1:A:663:CYS:SG	2.02	1.46
1:C:231:SER:HB2	4:C:801:HOH:O	1.45	1.13
1:D:261:VAL:HG13	4:D:809:HOH:O	1.68	0.92
1:D:306:ARG:HH11	1:D:306:ARG:HG3	1.40	0.87
1:C:663:CYS:SG	1:C:666:HIS:NE2	2.52	0.82

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the sym-

metry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:663:CYS:SG	1:C:663:CYS:SG[2_856]	1.53	0.67
1:B:663:CYS:SG	4:D:801:HOH:O[2_755]	2.13	0.07

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	414/504 (82%)	403 (97%)	9 (2%)	2 (0%)	29	22
1	B	416/504 (82%)	401 (96%)	14 (3%)	1 (0%)	47	45
1	C	421/504 (84%)	406 (96%)	14 (3%)	1 (0%)	47	45
1	D	421/504 (84%)	404 (96%)	16 (4%)	1 (0%)	47	45
2	E	5/11 (46%)	5 (100%)	0	0	100	100
2	F	6/11 (54%)	5 (83%)	1 (17%)	0	100	100
2	G	7/11 (64%)	6 (86%)	1 (14%)	0	100	100
2	H	6/11 (54%)	6 (100%)	0	0	100	100
All	All	1696/2060 (82%)	1636 (96%)	55 (3%)	5 (0%)	41	36

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	314	PRO
1	A	664	ARG
1	B	314	PRO
1	C	314	PRO
1	D	314	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	375/442 (85%)	365 (97%)	10 (3%)	44	43
1	B	377/442 (85%)	365 (97%)	12 (3%)	39	37
1	C	381/442 (86%)	368 (97%)	13 (3%)	37	34
1	D	381/442 (86%)	370 (97%)	11 (3%)	42	40
2	E	7/10 (70%)	7 (100%)	0	100	100
2	F	8/10 (80%)	8 (100%)	0	100	100
2	G	8/10 (80%)	8 (100%)	0	100	100
2	H	8/10 (80%)	8 (100%)	0	100	100
All	All	1545/1808 (86%)	1499 (97%)	46 (3%)	41	39

5 of 46 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	373	LEU
1	D	100	ASP
1	C	381	LYS
1	C	648	LYS
1	D	235	ARG

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

Mol	Chain	Res	Type
1	B	377	GLN
1	D	630	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 monosaccharides in this entry.

5.6 Ligand geometry [i](#)

4 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
3	NAG	G	200	2	14,14,15	3.44	8 (57%)	17,19,21	3.22	8 (47%)
3	NAG	H	200	2	14,14,15	3.26	4 (28%)	17,19,21	3.39	8 (47%)
3	NAG	F	200	2	14,14,15	3.34	5 (35%)	17,19,21	2.61	8 (47%)
3	NAG	E	200	2	14,14,15	3.86	7 (50%)	17,19,21	2.51	7 (41%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	NAG	G	200	2	-	1/6/23/26	0/1/1/1
3	NAG	H	200	2	-	2/6/23/26	0/1/1/1
3	NAG	F	200	2	-	2/6/23/26	0/1/1/1
3	NAG	E	200	2	-	2/6/23/26	0/1/1/1

The worst 5 of 24 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	E	200	NAG	C1-C2	11.89	1.70	1.52

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	F	200	NAG	C3-C2	9.26	1.72	1.52
3	H	200	NAG	C1-C2	8.18	1.64	1.52
3	H	200	NAG	C3-C2	7.16	1.67	1.52
3	G	200	NAG	C1-C2	6.81	1.62	1.52

The worst 5 of 31 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	H	200	NAG	C1-O5-C5	9.14	124.57	112.19
3	G	200	NAG	C1-O5-C5	8.07	123.13	112.19
3	G	200	NAG	O3-C3-C2	-5.70	97.68	109.47
3	F	200	NAG	C1-O5-C5	5.69	119.90	112.19
3	H	200	NAG	C3-C4-C5	-5.36	100.67	110.24

There are no chirality outliers.

5 of 7 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	H	200	NAG	C1-C2-N2-C7
3	E	200	NAG	C1-C2-N2-C7
3	G	200	NAG	C3-C2-N2-C7
3	E	200	NAG	C3-C2-N2-C7
3	F	200	NAG	C3-C2-N2-C7

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	424/504 (84%)	-0.25	3 (0%) 87 90	10, 25, 47, 71	0
1	B	426/504 (84%)	-0.19	5 (1%) 79 83	10, 26, 51, 74	0
1	C	429/504 (85%)	-0.26	0 100 100	12, 27, 51, 70	0
1	D	429/504 (85%)	-0.20	3 (0%) 87 90	12, 29, 54, 75	0
2	E	7/11 (63%)	-0.57	0 100 100	22, 28, 31, 34	0
2	F	8/11 (72%)	-0.23	0 100 100	23, 29, 38, 54	0
2	G	9/11 (81%)	-0.27	0 100 100	24, 30, 38, 50	0
2	H	8/11 (72%)	-0.36	0 100 100	26, 34, 39, 39	0
All	All	1740/2060 (84%)	-0.23	11 (0%) 89 91	10, 27, 51, 75	0

The worst 5 of 11 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	604	ILE	3.6
1	A	605	GLU	3.2
1	A	553	VAL	3.2
1	A	664	ARG	3.0
1	D	694	PRO	2.6

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
3	NAG	E	200	14/15	0.98	0.12	12,16,18,19	0
3	NAG	F	200	14/15	0.98	0.13	12,16,21,22	0
3	NAG	G	200	14/15	0.98	0.12	13,15,20,21	0
3	NAG	H	200	14/15	0.98	0.13	15,17,23,24	0

6.5 Other polymers [i](#)

There are no such residues in this entry.