wwPDB X-ray Structure Validation Summary Report (i)

Feb 18, 2024 – 09:41 PM EST

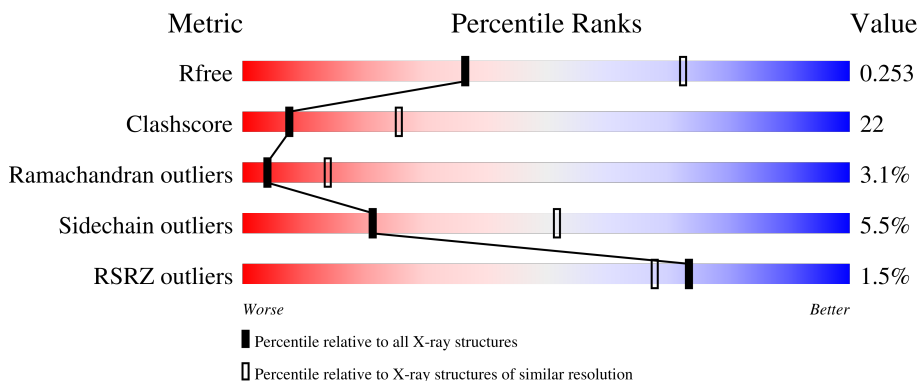|		
PDB ID	:	4GDF
Title	:	A Crystal Structure of SV40 Large T Antigen
Authors	:	Chang, Y.P.; Xu, M.; Chen, X.S.
Deposited on	:	2012-07-31
Resolution	:	2.80 Å(reported)

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.
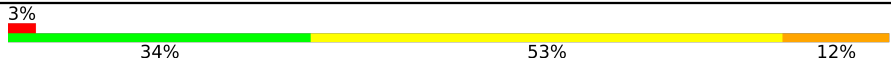

Metric	Percentile Ranks	Value
Rfree		0.253
Clashscore		22
Ramachandran outliers		3.1%
Sidechain outliers		5.5%
RSRZ outliers		1.5%

Worse ■ Percentile relative to all X-ray structures
☐ Percentile relative to X-ray structures of similar resolution *Better*
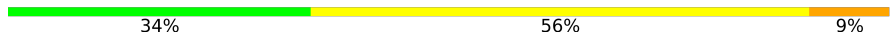
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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 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	497	% 58% 37% 5%
1	B	497	% 60% 36% •
1	E	497	% 58% 37% •
1	F	497	2% 59% 38% •
2	C	32	3% 34% 53% 12%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	G	32	3% 34% 53% 12%
3	D	32	34% 56% 9%
3	H	32	34% 56% 9%

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 18835 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 Large T antigen.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	497	Total 4035	C 2601	N 675	O 733	S 26	0	0	0
1	B	497	Total 4035	C 2601	N 675	O 733	S 26	0	0	0
1	E	497	Total 4035	C 2601	N 675	O 733	S 26	0	0	0
1	F	497	Total 4035	C 2601	N 675	O 733	S 26	0	0	0

- Molecule 2 is a DNA chain called DNA (32-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	32	Total 657	C 312	N 126	O 188	P 31	0	0	0
2	G	32	Total 657	C 312	N 126	O 188	P 31	0	0	0

- Molecule 3 is a DNA chain called DNA (32-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	32	Total 649	C 310	N 116	O 192	P 31	0	0	0
3	H	32	Total 649	C 310	N 116	O 192	P 31	0	0	0

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	E	1	Total 1	Zn 1	0	0
4	F	1	Total 1	Zn 1	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	18	Total 18	O 18	0	0
5	B	12	Total 12	O 12	0	0
5	C	9	Total 9	O 9	0	0
5	D	6	Total 6	O 6	0	0
5	E	14	Total 14	O 14	0	0
5	F	7	Total 7	O 7	0	0
5	G	6	Total 6	O 6	0	0
5	H	7	Total 7	O 7	0	0

3 Residue-property plots ⓘ

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: Large T antigen

- Molecule 1: Large T antigen

T433 T434 L435 L440 G444 G445 L448 N449 V450 N451 L452 P453 L454 D455 R456 L457 E460 L461 G462 V463 D466 Q467 F468 L469 V470 V471 D474 V475 K476 G480 R483 D484 L485 I491 R498 E510 H513 L514 R517 I520 P523 G524 I525 M528

V533 P534 K535 T536 L537 Q538 A539 R540 F541 V542 D546 F547 L553 K554 H555 C556 L557 E558 R559 F562 L563 R567 I568 I569 Q570 I573 L576 L577 M578 L579 I580 W581 Y582 R583 P584 V585 A586 E587 F588 A589 Q590 S591 I592 Q593 S594 R595 I596 V597 E598 W599 K600 L603

F607 S608 L609 S610 V611 M615 K616 F617 N618 I623 G624 V625 L626 D627

● **Molecule 1: Large T antigen**

Chain E:

K131 V132 E133 D134 P135 K136 D137 S140 L143 S144 F145 A149 V150 F151 R154 T155 L156 A160 I161 Y162 T163 L170 K173 K178 V181 T182 I183 I184 S185 R186 N193 I194 L195 F196 F197 L198 T199 P200 H201 R202 H203 R204 V205 S206 N209 N210 Y211 A212 Q213

K214 L215 C216 T217 F218 S219 F220 L221 I222 C223 K224 G225 K228 E229 Y230 L231 M232 Y233 D239 P240 F241 S242 V243 I244 E246 S247 L248 P249 G250 Q251 L252 K253 E254 H255 D256 F257 M258 P259 E260 E261 A262 K266 W270 D283 D284 V285 M291 E294 F299 E300 M301

C302 L303 K304 C305 K308 P311 S312 H313 Y314 E318 A324 A325 I326 F327 A328 K331 I336 C337 Q338 V341 R349 T356 R357 E358 Q359 N363 N366 L369 M372 M375 F376 G377 A382 E385 M388 A389 G390 V391 L394 L397

M401 L409 V413 Y414 N415 K418 K419 R420 Y421 W422 L423 F424 I428 A437 L440 L452 P453 L454 E460 V463 D466 Q467 F468 L469 V470 V471 F472 V475 K476 G477 T478 G479 G480 S482 R483 D484 S487 I491 L494 R498 D499 D502 K506 K512 H513 K516 R517 T518 Q519 I520 F521 P522 P523 G524 I525 V526 T527 M528 N529 E530 V533 L537 R540 F541 V542 K543 Q544 L553 R559 S560 I563 L564 E565 K566 R567 Q570 L575 L576 L577 M578 L579 I580 R583 E587 F588

A589 Q590 S591 I592 Q593 S594 R595 I596 V597 E598 W599 K600 E601 R602 L603 D604 K605 E606 F607 S608 L609 S610 V611 Y612 M615 K616 F617 N618 I623 G624 V625 L626 D627

● **Molecule 1: Large T antigen**

Chain F:

K131 V132 E133 D134 P135 K136 D137 S140 L143 S144 F145 H148 A149 V150 R154 T155 L156 A157 C158 F159 A160 I161 Y162 T163 E166 M176 V181 I184 H187 N188 H192 L195 F196 F197 R202 H203 R204 N209 Q213 F220 L221 I222 C223 K224 G225

V226 N227 K228 L231 M232 Y233 L236 I244 S247 G251 L252 F257 N258 P259 E263 E264 W270 K271 T274 E275 M278 K281 C282 D283 D284 V285 L286 L287 M291 E294 F299 E300 M301 K304 E309 Q310 P311 S312 H313 E318 K331

I336 C337 Q338 Q339 A340 V341 D342 T343 V344 R349 V350 D351 S352 L353 Q354 L355 T356 R357 E358 Q359 M360 R364 L368 L369 D370 R371 M372 M375 I384 E385 E386 W387 M388 A389 G390 V391 A392 W393 L394 H395 C396 L397 M401 D402 Y406 D407 F408 L409 K410 C411 M412 V413 Y414

N415 I416 K419 R420 L423 K432 T433 T434 L435 L439 L440 G444 G445 L448 N449 V450 N451 L452 P453 L454 D455 R456 L457 E460 L461 G462 D466 Q467 F468 L469 V470 V471 D474 V475 K476 G480 R483 D484 L485 I491 R498 E510 H513 L514

R517 I520 P523 G524 I525 M528 V533 P534 K535 T536 L537 Q538 A539 R540 F541 V542 D546 F547 L553 K554 H555 C556 L557 E558 R559 F562 L563 R567 I568 I569 Q570 I573 L576 L577 M578 L579 I580 W581 Y582 R583 P584 V585 A586 E587 F588 A589 Q590 S591 I592 Q593

S594 R595 I596 V597 E598 W599 K600 L603 F607 S608 L609 S610 V611 M615 K616 F617 N618 I623 G624 V625 L626 D627

• Molecule 2: DNA (32-MER)

Chain C:

A1 C2 T3 A4 C5 T6 T7 C8 T9 G10 G11 A12 A13 G16 C17 T18 C19 A20 G21 A22 G23 A28 G29 G32

• Molecule 2: DNA (32-MER)

Chain G:

A1 C2 T3 A4 C5 T6 T7 C8 T9 G10 G11 A12 A13 G16 C17 T18 C19 A20 G21 A22 G23 A28 G29 G32

• Molecule 3: DNA (32-MER)

Chain D:

C1 G2 C3 C4 T5 C6 G7 G8 C12 T13 G14 A15 G16 C17 T18 A19 T20 T21 C22 C23 A24 G25 G31 T32

• Molecule 3: DNA (32-MER)

Chain H:

C1 G2 C3 C4 T5 C6 G7 G8 C12 T13 G14 A15 G16 C17 T18 A19 T20 T21 C22 C23 A24 G25 G31 T32

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	73.63Å 128.31Å 166.12Å 90.00° 89.91° 90.00°	Depositor
Resolution (Å)	50.00 – 2.80 48.37 – 2.81	Depositor EDS
% Data completeness (in resolution range)	(Not available) (50.00-2.80) 86.5 (48.37-2.81)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.73 (at 2.81Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.227 , 0.255 0.224 , 0.253	Depositor DCC
R_{free} test set	3576 reflections (5.10%)	wwPDB-VP
Wilson B-factor (Å^2)	51.4	Xtriage
Anisotropy	0.376	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 26.3	EDS
L-test for twinning[2]	$<|L|> = 0.48, <L^2> = 0.31$	Xtriage
Estimated twinning fraction	0.460 for h,-k,-l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	18835	wwPDB-VP
Average B, all atoms (Å^2)	63.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.16% 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 ⓘ

5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	0.47	0/4124	0.63	3/5562 (0.1%)
1	B	0.45	0/4124	0.62	0/5562
1	E	0.47	0/4124	0.63	3/5562 (0.1%)
1	F	0.45	0/4124	0.62	0/5562
2	C	0.54	0/738	1.00	5/1138 (0.4%)
2	G	0.55	0/738	1.02	5/1138 (0.4%)
3	D	0.54	0/726	0.95	1/1118 (0.1%)
3	H	0.55	0/726	0.96	1/1118 (0.1%)
All	All	0.48	0/19424	0.70	18/26760 (0.1%)

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	B	0	1
1	F	0	1
2	C	0	2
2	G	0	2
3	D	0	3
3	H	0	3
All	All	0	12

There are no bond length outliers.

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	G	1	DA	O4'-C1'-N9	-12.03	99.58	108.00
2	C	3	DT	N1-C1'-C2'	10.90	133.31	112.60

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	G	3	DT	N1-C1'-C2'	10.71	132.94	112.60
2	C	1	DA	O4'-C1'-N9	-10.46	100.68	108.00
3	D	31	DG	N9-C1'-C2'	8.13	128.06	112.60

There are no chirality outliers.

5 of 12 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	309	GLU	Mainchain
2	C	16	DG	Sidechain
2	C	17	DC	Sidechain
3	D	5	DT	Sidechain
3	D	8	DG	Sidechain

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	4035	0	4068	179	0
1	B	4035	0	4068	180	0
1	E	4035	0	4068	177	0
1	F	4035	0	4068	182	0
2	C	657	0	360	31	0
2	G	657	0	360	29	0
3	D	649	0	362	26	0
3	H	649	0	362	26	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
4	E	1	0	0	0	0
4	F	1	0	0	0	0
5	A	18	0	0	0	0
5	B	12	0	0	1	0
5	C	9	0	0	0	0
5	D	6	0	0	0	0
5	E	14	0	0	1	0
5	F	7	0	0	0	0
5	G	6	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	H	7	0	0	0	0
All	All	18835	0	17716	798	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 22.

The worst 5 of 798 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:G:1:DA:H2'	2:G:2:DC:C5	1.63	1.31
2:C:1:DA:H2'	2:C:2:DC:C5	1.64	1.30
2:G:1:DA:C2'	2:G:2:DC:C5	2.21	1.23
2:C:1:DA:C2'	2:C:2:DC:C5	2.21	1.22
1:F:498:ARG:HD2	1:F:540:ARG:HH21	1.11	1.09

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	495/497 (100%)	423 (86%)	52 (10%)	20 (4%)	3	9
1	B	495/497 (100%)	416 (84%)	68 (14%)	11 (2%)	6	22
1	E	495/497 (100%)	421 (85%)	55 (11%)	19 (4%)	3	10
1	F	495/497 (100%)	417 (84%)	67 (14%)	11 (2%)	6	22
All	All	1980/1988 (100%)	1677 (85%)	242 (12%)	61 (3%)	4	14

5 of 61 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	258	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	357	ARG
1	A	479	GLY
1	A	480	GLY
1	A	481	GLU

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	446/446 (100%)	422 (95%)	24 (5%)	22	53
1	B	446/446 (100%)	420 (94%)	26 (6%)	20	50
1	E	446/446 (100%)	423 (95%)	23 (5%)	23	55
1	F	446/446 (100%)	421 (94%)	25 (6%)	21	51
All	All	1784/1784 (100%)	1686 (94%)	98 (6%)	21	52

5 of 98 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	443	CYS
1	E	615	MET
1	E	452	LEU
1	E	529	ASN
1	F	145	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 46 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	519	GLN
1	F	267	GLN
1	E	529	ASN
1	F	209	ASN
1	F	338	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 monosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 4 ligands modelled in this entry, 4 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 ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

6 Fit of model and data ⓘ

6.1 Protein, DNA and RNA chains ⓘ

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(Å²)	Q<0.9
1	A	497/497 (100%)	-0.19	6 (1%)	79	73	34, 56, 97, 138	0
1	B	497/497 (100%)	-0.06	7 (1%)	75	70	29, 59, 107, 126	0
1	E	497/497 (100%)	-0.15	5 (1%)	82	77	33, 56, 96, 138	0
1	F	497/497 (100%)	-0.07	11 (2%)	62	52	29, 59, 107, 126	0
2	C	32/32 (100%)	-0.30	1 (3%)	49	39	29, 55, 160, 182	0
2	G	32/32 (100%)	-0.46	1 (3%)	49	39	30, 55, 160, 182	0
3	D	32/32 (100%)	-0.41	0	100	100	30, 48, 107, 139	0
3	H	32/32 (100%)	-0.51	0	100	100	32, 48, 107, 140	0
All	All	2116/2116 (100%)	-0.13	31 (1%)	73	68	29, 58, 106, 182	0

The worst 5 of 31 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	259	PRO	4.9
1	B	257	PHE	4.7
1	B	263	GLU	4.4
2	C	1	DA	4.2
1	B	301	MET	3.9

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

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

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

6.4 Ligands ⓘ

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
4	ZN	F	700	1/1	0.94	0.14	69,69,69,69	0
4	ZN	B	700	1/1	0.95	0.13	68,68,68,68	0
4	ZN	E	700	1/1	0.97	0.14	58,58,58,58	0
4	ZN	A	700	1/1	0.99	0.16	58,58,58,58	0

6.5 Other polymers ⓘ

There are no such residues in this entry.