



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 19, 2016 – 08:52 PM GMT

PDB ID : 4YWT
Title : Crystal structure of full-length glypican-1 core protein after controlled crystal dehydration to 87% relative humidity
Authors : Awad, W.; Mani, K.; Logan, D.T.
Deposited on : 2015-03-21
Resolution : 2.38 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
<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 : 1.7.1 (RC1), CSD as537be (2016)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026982
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) : rb-20026982

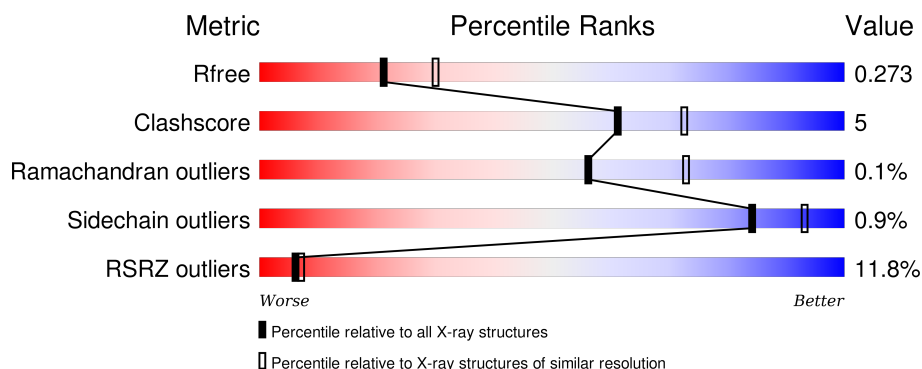
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.38 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	4019 (2.40-2.36)
Clashscore	102246	4595 (2.40-2.36)
Ramachandran outliers	100387	4520 (2.40-2.36)
Sidechain outliers	100360	4522 (2.40-2.36)
RSRZ outliers	91569	4034 (2.40-2.36)

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. 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	526	<div> <div>11%</div> <div>69%</div> <div>9%</div> <div>22%</div> </div>
1	B	526	<div> <div>9%</div> <div>70%</div> <div>12%</div> <div>18%</div> </div>
1	C	526	<div> <div>6%</div> <div>68%</div> <div>8%</div> <div>24%</div> </div>
1	D	526	<div> <div>11%</div> <div>67%</div> <div>15%</div> <div>18%</div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	CA	C	603	-	-	-	X

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 13110 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 Glypican-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	411	Total	C	N	O	S	0	4	0
			3166	1981	568	594	23			
1	B	431	Total	C	N	O	S	0	3	0
			3283	2062	581	616	24			
1	C	400	Total	C	N	O	S	0	2	0
			3058	1918	544	575	21			
1	D	429	Total	C	N	O	S	0	4	0
			3254	2043	571	613	27			

There are 100 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	2	ALA	-	expression tag	UNP P35052
A	3	PRO	-	expression tag	UNP P35052
A	4	GLN	-	expression tag	UNP P35052
A	5	LEU	-	expression tag	UNP P35052
A	6	HIS	-	expression tag	UNP P35052
A	7	HIS	-	expression tag	UNP P35052
A	8	HIS	-	expression tag	UNP P35052
A	9	HIS	-	expression tag	UNP P35052
A	10	HIS	-	expression tag	UNP P35052
A	11	HIS	-	expression tag	UNP P35052
A	12	ASP	-	expression tag	UNP P35052
A	13	LEU	-	expression tag	UNP P35052
A	14	TYR	-	expression tag	UNP P35052
A	15	GLU	-	expression tag	UNP P35052
A	16	ASN	-	expression tag	UNP P35052
A	17	LEU	-	expression tag	UNP P35052
A	18	TYR	-	expression tag	UNP P35052
A	19	PHE	-	expression tag	UNP P35052
A	20	GLN	-	expression tag	UNP P35052
A	21	GLY	-	expression tag	UNP P35052
A	22	LYS	-	expression tag	UNP P35052

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	23	LEU	-	expression tag	UNP P35052
A	486	ALA	SER	engineered mutation	UNP P35052
A	488	ALA	SER	engineered mutation	UNP P35052
A	490	ALA	SER	engineered mutation	UNP P35052
B	2	ALA	-	expression tag	UNP P35052
B	3	PRO	-	expression tag	UNP P35052
B	4	GLN	-	expression tag	UNP P35052
B	5	LEU	-	expression tag	UNP P35052
B	6	HIS	-	expression tag	UNP P35052
B	7	HIS	-	expression tag	UNP P35052
B	8	HIS	-	expression tag	UNP P35052
B	9	HIS	-	expression tag	UNP P35052
B	10	HIS	-	expression tag	UNP P35052
B	11	HIS	-	expression tag	UNP P35052
B	12	ASP	-	expression tag	UNP P35052
B	13	LEU	-	expression tag	UNP P35052
B	14	TYR	-	expression tag	UNP P35052
B	15	GLU	-	expression tag	UNP P35052
B	16	ASN	-	expression tag	UNP P35052
B	17	LEU	-	expression tag	UNP P35052
B	18	TYR	-	expression tag	UNP P35052
B	19	PHE	-	expression tag	UNP P35052
B	20	GLN	-	expression tag	UNP P35052
B	21	GLY	-	expression tag	UNP P35052
B	22	LYS	-	expression tag	UNP P35052
B	23	LEU	-	expression tag	UNP P35052
B	486	ALA	SER	engineered mutation	UNP P35052
B	488	ALA	SER	engineered mutation	UNP P35052
B	490	ALA	SER	engineered mutation	UNP P35052
C	2	ALA	-	expression tag	UNP P35052
C	3	PRO	-	expression tag	UNP P35052
C	4	GLN	-	expression tag	UNP P35052
C	5	LEU	-	expression tag	UNP P35052
C	6	HIS	-	expression tag	UNP P35052
C	7	HIS	-	expression tag	UNP P35052
C	8	HIS	-	expression tag	UNP P35052
C	9	HIS	-	expression tag	UNP P35052
C	10	HIS	-	expression tag	UNP P35052
C	11	HIS	-	expression tag	UNP P35052
C	12	ASP	-	expression tag	UNP P35052
C	13	LEU	-	expression tag	UNP P35052
C	14	TYR	-	expression tag	UNP P35052

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	15	GLU	-	expression tag	UNP P35052
C	16	ASN	-	expression tag	UNP P35052
C	17	LEU	-	expression tag	UNP P35052
C	18	TYR	-	expression tag	UNP P35052
C	19	PHE	-	expression tag	UNP P35052
C	20	GLN	-	expression tag	UNP P35052
C	21	GLY	-	expression tag	UNP P35052
C	22	LYS	-	expression tag	UNP P35052
C	23	LEU	-	expression tag	UNP P35052
C	486	ALA	SER	engineered mutation	UNP P35052
C	488	ALA	SER	engineered mutation	UNP P35052
C	490	ALA	SER	engineered mutation	UNP P35052
D	2	ALA	-	expression tag	UNP P35052
D	3	PRO	-	expression tag	UNP P35052
D	4	GLN	-	expression tag	UNP P35052
D	5	LEU	-	expression tag	UNP P35052
D	6	HIS	-	expression tag	UNP P35052
D	7	HIS	-	expression tag	UNP P35052
D	8	HIS	-	expression tag	UNP P35052
D	9	HIS	-	expression tag	UNP P35052
D	10	HIS	-	expression tag	UNP P35052
D	11	HIS	-	expression tag	UNP P35052
D	12	ASP	-	expression tag	UNP P35052
D	13	LEU	-	expression tag	UNP P35052
D	14	TYR	-	expression tag	UNP P35052
D	15	GLU	-	expression tag	UNP P35052
D	16	ASN	-	expression tag	UNP P35052
D	17	LEU	-	expression tag	UNP P35052
D	18	TYR	-	expression tag	UNP P35052
D	19	PHE	-	expression tag	UNP P35052
D	20	GLN	-	expression tag	UNP P35052
D	21	GLY	-	expression tag	UNP P35052
D	22	LYS	-	expression tag	UNP P35052
D	23	LEU	-	expression tag	UNP P35052
D	486	ALA	SER	engineered mutation	UNP P35052
D	488	ALA	SER	engineered mutation	UNP P35052
D	490	ALA	SER	engineered mutation	UNP P35052

- Molecule 2 is N-ACETYL-D-GLUCOSAMINE (three-letter code: NAG) (formula: C₈H₁₅NO₆).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	0	0
			14	8	1	5		
2	B	1	Total	C	N	O	0	0
			14	8	1	5		
2	C	1	Total	C	N	O	0	0
			14	8	1	5		
2	D	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	2	Total	Ca	0	0
			2	2		
3	A	2	Total	Ca	0	0
			2	2		
3	D	2	Total	Ca	0	0
			2	2		
3	C	3	Total	Ca	0	0
			3	3		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	102	Total	O	0	0
			102	102		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	86	Total 86	O 86	0	0
4	C	53	Total 53	O 53	0	0
4	D	43	Total 43	O 43	0	0

[illegible][illegible]

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	46.78 Å 166.59 Å 137.70 Å 90.00° 90.38° 90.00°	Depositor
Resolution (Å)	43.22 – 2.38 43.22 – 2.38	Depositor EDS
% Data completeness (in resolution range)	97.8 (43.22-2.38) 97.9 (43.22-2.38)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.35 (at 2.37 Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, R_{free}	0.239 , 0.273 0.239 , 0.273	Depositor DCC
R_{free} test set	1977 reflections (2.40%)	DCC
Wilson B-factor (Å ²)	38.7	Xtriage
Anisotropy	0.569	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 59.8	EDS
Estimated twinning fraction	0.055 for h,-k,-l	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	1 of 82475 reflections (0.001%)	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	13110	wwPDB-VP
Average B, all atoms (Å ²)	59.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 46.82 % 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 1.0833e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹ Intensities estimated from amplitudes.

² 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: CA, 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	0.24	0/3238	0.45	1/4390 (0.0%)
1	B	0.29	0/3352	0.47	1/4550 (0.0%)
1	C	0.26	0/3120	0.44	0/4234
1	D	0.25	0/3326	0.44	0/4517
All	All	0.26	0/13036	0.45	2/17691 (0.0%)

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	A	0	1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	A	336	THR	N-CA-C	6.11	127.49	111.00
1	B	404	LYS	CA-CB-CG	5.15	124.74	113.40

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	198	ALA	Peptide

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	3166	0	3070	30	0
1	B	3283	0	3184	39	0
1	C	3058	0	2949	25	0
1	D	3254	0	3126	45	0
2	A	14	0	13	0	0
2	B	14	0	13	0	0
2	C	14	0	13	0	0
2	D	14	0	13	0	0
3	A	2	0	0	0	0
3	B	2	0	0	0	0
3	C	3	0	0	0	0
3	D	2	0	0	0	0
4	A	102	0	0	3	0
4	B	86	0	0	4	0
4	C	53	0	0	0	0
4	D	43	0	0	2	0
All	All	13110	0	12381	138	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.

The worst 5 of 138 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:407:LEU:HD22	1:B:415:CYS:HB3	1.66	0.77
1:A:25:PRO:HD2	1:A:261:GLY:HA2	1.67	0.76
1:B:319:THR:HG21	1:B:367:ARG:HA	1.69	0.74
1:B:268[B]:CYS:SG	4:B:781:HOH:O	2.47	0.73
1:B:129:ALA:HB2	1:B:364:LEU:HD11	1.70	0.72

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	407/526 (77%)	401 (98%)	4 (1%)	2 (0%)	34	46
1	B	428/526 (81%)	412 (96%)	16 (4%)	0	100	100
1	C	394/526 (75%)	385 (98%)	9 (2%)	0	100	100
1	D	427/526 (81%)	418 (98%)	9 (2%)	0	100	100
All	All	1656/2104 (79%)	1616 (98%)	38 (2%)	2 (0%)	56	73

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	336	THR
1	A	337	ALA

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	332/439 (76%)	331 (100%)	1 (0%)	94	98
1	B	341/439 (78%)	336 (98%)	5 (2%)	72	86
1	C	317/439 (72%)	315 (99%)	2 (1%)	90	96
1	D	336/439 (76%)	331 (98%)	5 (2%)	72	86
All	All	1326/1756 (76%)	1313 (99%)	13 (1%)	84	92

5 of 13 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	474	ASN
1	C	94	ARG
1	D	165	GLU
1	B	405	MET
1	D	136	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	181	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 13 ligands modelled in this entry, 9 are monoatomic - leaving 4 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	$\# Z > 2$
2	NAG	A	601	1	14,14,15	0.39	0	15,19,21	0.45	0
2	NAG	B	601	1	14,14,15	0.30	0	15,19,21	0.32	0
2	NAG	C	601	1	14,14,15	0.33	0	15,19,21	0.45	0
2	NAG	D	601	1	14,14,15	0.23	0	15,19,21	0.50	0

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
2	NAG	A	601	1	-	0/6/23/26	0/1/1/1
2	NAG	B	601	1	-	0/6/23/26	0/1/1/1
2	NAG	C	601	1	-	0/6/23/26	0/1/1/1
2	NAG	D	601	1	-	0/6/23/26	0/1/1/1

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 [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	411/526 (78%)	1.01	58 (14%) 4 4	25, 48, 96, 119	0
1	B	431/526 (81%)	0.94	48 (11%) 7 8	30, 53, 98, 134	0
1	C	400/526 (76%)	0.89	33 (8%) 14 16	40, 59, 107, 139	0
1	D	429/526 (81%)	0.99	59 (13%) 4 5	37, 60, 104, 135	0
All	All	1671/2104 (79%)	0.96	198 (11%) 6 7	25, 56, 102, 139	0

The worst 5 of 198 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	180	PRO	9.3
1	A	196	ALA	8.3
1	C	337	ALA	8.1
1	B	156	ASN	8.1
1	A	184	LEU	6.3

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, 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	LLDF	B-factors(\AA^2)	Q<0.9
3	CA	C	603	1/1	0.34	0.27	3.58	99,99,99,99	0
3	CA	A	603	1/1	0.27	0.15	-0.40	90,90,90,90	0
3	CA	C	604	1/1	0.43	0.18	-0.78	60,60,60,60	0
2	NAG	A	601	14/15	0.90	0.16	-1.31	56,68,78,81	0
3	CA	D	602	1/1	0.39	0.11	-1.71	93,93,93,93	0
3	CA	B	602	1/1	0.91	0.06	-4.78	87,87,87,87	0
2	NAG	D	601	14/15	0.86	0.20	-	80,89,95,96	0
3	CA	A	602	1/1	0.70	0.15	-	47,47,47,47	0
3	CA	B	603	1/1	0.28	0.10	-	43,43,43,43	1
2	NAG	C	601	14/15	0.86	0.20	-	79,84,89,90	0
3	CA	C	602	1/1	0.72	0.14	-	71,71,71,71	0
2	NAG	B	601	14/15	0.87	0.19	-	65,72,80,84	0
3	CA	D	603	1/1	0.86	0.11	-	82,82,82,82	0

6.5 Other polymers [i](#)

There are no such residues in this entry.