



# Full wwPDB X-ray Structure Validation Report ⓘ

Nov 6, 2023 – 11:43 PM EST

PDB ID : 4KA2  
Title : Crystal structure of CD4-mimetic miniprotein M48U12 in complex with HIV-1 YU2 gp120  
Authors : Acharya, P.; Kwong, P.D.  
Deposited on : 2013-04-21  
Resolution : 1.79 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](mailto: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 : 2.36  
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) : 2.36

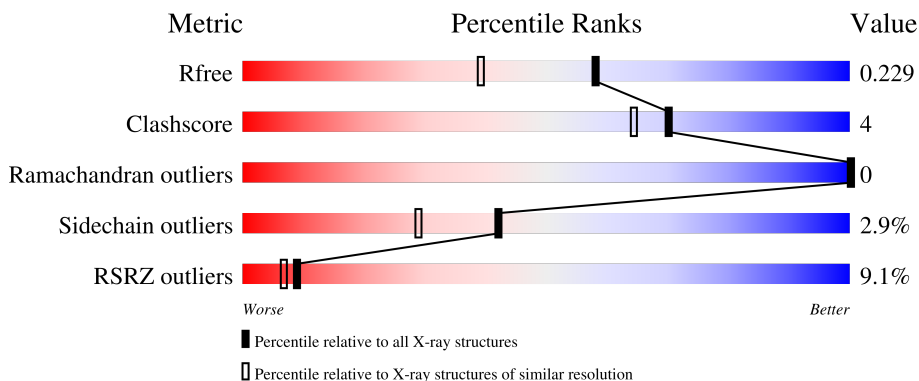
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*


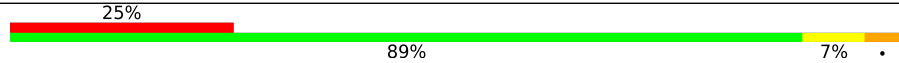
The reported resolution of this entry is 1.79 Å.

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	5950 (1.80-1.80)
Clashscore	141614	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)
RSRZ outliers	127900	5850 (1.80-1.80)

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  $\geq 3$ , 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	376	
2	R	28	

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 6202 atoms, of which 2911 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.

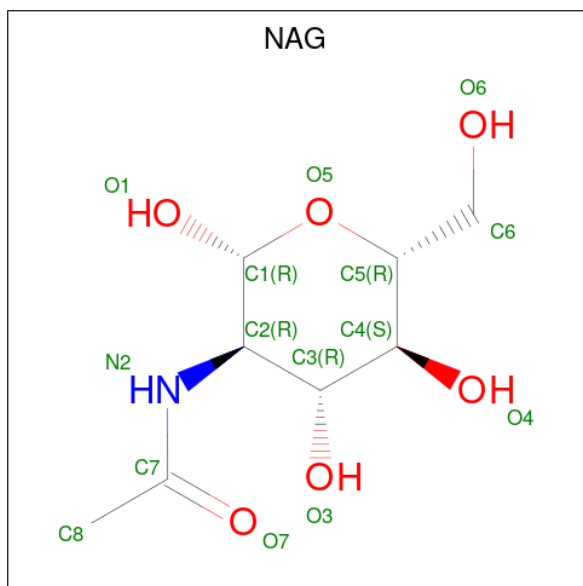
- Molecule 1 is a protein called HIV-1 YU2 gp120.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	H	N	O				S
1	A	337	5371	1705	2662	460	515	29	0	21	0

- Molecule 2 is a protein called M48U12.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	H	N	O				S
2	R	28	483	154	249	42	32	6	0	3	1

- Molecule 3 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula:  $C_8H_{15}NO_6$ ).



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

*Continued on next page...*

*Continued from previous page...*

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

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	227	Total	O	0	0
			227	227		
4	R	9	Total	O	0	0
			9	9		



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	64.37Å 163.44Å 78.17Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	17.48 – 1.79 17.48 – 1.79	Depositor EDS
% Data completeness (in resolution range)	81.7 (17.48-1.79) 75.2 (17.48-1.79)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.74 (at 1.79Å)	Xtrriage
Refinement program	PHENIX (phenix.refine: 1.8.2_1309)	Depositor
R, $R_{free}$	0.167 , 0.228 0.169 , 0.229	Depositor DCC
$R_{free}$ test set	2000 reflections (6.24%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	23.1	Xtrriage
Anisotropy	0.423	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.39 , 49.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	6202	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	46.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.39% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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: MPT, DPR, U3X, NH2, 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	0.60	0/2838	0.67	0/3853
2	R	0.51	0/191	0.69	0/250
All	All	0.59	0/3029	0.67	0/4103

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.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	232	LYS	Peptide

### 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	2709	2662	2664	19	0
2	R	234	249	243	3	0
3	A	112	0	104	4	0
4	A	227	0	0	7	0
4	R	9	0	0	0	0
All	All	3291	2911	3011	22	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 4.

All (22) 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:R:1:MPT:SG	2:R:1:MPT:CB	2.11	1.39
1:A:247[B]:CYS:SG	4:A:664:HOH:O	2.38	0.81
1:A:205[B]:CYS:SG	4:A:765:HOH:O	2.38	0.80
1:A:247[B]:CYS:SG	4:A:764:HOH:O	2.39	0.80
1:A:82:GLN:NE2	4:A:689:HOH:O	2.26	0.68
1:A:106[A]:GLU:OE1	4:A:708:HOH:O	2.10	0.68
1:A:124:GLY:O	4:A:747:HOH:O	2.14	0.65
3:A:502:NAG:H82	3:A:502:NAG:O3	1.99	0.62
1:A:54[B]:CYS:HB2	1:A:74:CYS:SG	2.42	0.58
2:R:1:MPT:SG	2:R:1:MPT:CA	2.92	0.56
1:A:419:ARG:NH1	4:A:721:HOH:O	2.36	0.53
1:A:54[B]:CYS:SG	1:A:217:TYR:CE1	3.00	0.50
1:A:289:ASN:ND2	3:A:505:NAG:O7	2.47	0.47
1:A:272:ILE:HG22	1:A:286:VAL:HG22	1.97	0.47
1:A:211[B]:GLU:OE2	3:A:504:NAG:H4	2.14	0.47
1:A:93:PHE:HB2	1:A:233:PHE:HZ	1.83	0.44
1:A:378:CYS:SG	1:A:445[B]:CYS:HB3	2.57	0.44
1:A:333:LEU:N	1:A:333:LEU:HD23	2.33	0.44
1:A:56:SER:C	1:A:77:THR:HG23	2.38	0.43
1:A:382:PHE:HD1	2:R:23[B]:U3X:H14	1.82	0.43
1:A:86:LEU:HD11	1:A:244[B]:THR:CG2	2.52	0.40
1:A:85:LYS:NZ	3:A:503:NAG:HN2	2.19	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	351/376 (93%)	343 (98%)	8 (2%)	0	100	100
2	R	26/28 (93%)	24 (92%)	2 (8%)	0	100	100
All	All	377/404 (93%)	367 (97%)	10 (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	316/328 (96%)	307 (97%)	9 (3%)	43	30
2	R	22/20 (110%)	21 (96%)	1 (4%)	27	13
All	All	338/348 (97%)	328 (97%)	10 (3%)	42	27

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	45	TRP
1	A	57	ASP
1	A	82	GLN
1	A	85	LYS
1	A	201[A]	ILE
1	A	201[B]	ILE
1	A	226	LEU
1	A	333	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	395	ASP
2	R	13	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

3 non-standard protein/DNA/RNA residues 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	U3X	R	23[B]	-	19,20,21	0.99	1 (5%)	22,25,27	1.13	2 (9%)
2	U3X	R	23[A]	-	19,20,21	1.10	1 (5%)	22,25,27	1.15	2 (9%)

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	U3X	R	23[B]	-	-	2/10/19/21	0/2/2/2
2	U3X	R	23[A]	-	-	5/10/19/21	0/2/2/2

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	R	23[A]	U3X	C06-N01	3.11	1.47	1.38
2	R	23[B]	U3X	C06-N01	2.58	1.46	1.38

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	R	23[A]	U3X	C13-C08-C07	3.10	118.07	111.47
2	R	23[B]	U3X	CB-CA-C	-2.84	106.14	111.47
2	R	23[A]	U3X	CB-CA-C	-2.16	107.42	111.47
2	R	23[B]	U3X	C07-N01-C06	-2.11	116.69	122.14

There are no chirality outliers.

All (7) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	R	23[A]	U3X	N01-C07-C08-C13
2	R	23[B]	U3X	C04-C03-CB-CA
2	R	23[B]	U3X	C02-C03-CB-CA
2	R	23[A]	U3X	N01-C07-C08-C09
2	R	23[A]	U3X	C02-C03-CB-CA
2	R	23[A]	U3X	C04-C03-CB-CA
2	R	23[A]	U3X	C08-C07-N01-C06

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	R	23[B]	U3X	1	0

## 5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

8 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	A	506	1	14,14,15	0.19	0	17,19,21	0.51	0
3	NAG	A	502	1	14,14,15	0.69	1 (7%)	17,19,21	0.48	0
3	NAG	A	507	1	14,14,15	0.25	0	17,19,21	0.64	0
3	NAG	A	505	1	14,14,15	0.36	0	17,19,21	0.63	0
3	NAG	A	503	1	14,14,15	0.75	1 (7%)	17,19,21	0.45	0
3	NAG	A	508	1	14,14,15	0.38	0	17,19,21	0.54	0
3	NAG	A	504	1	14,14,15	0.65	0	17,19,21	0.66	0
3	NAG	A	501	1	14,14,15	0.66	0	17,19,21	0.53	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
3	NAG	A	506	1	-	0/6/23/26	0/1/1/1
3	NAG	A	502	1	-	2/6/23/26	0/1/1/1
3	NAG	A	507	1	-	0/6/23/26	0/1/1/1
3	NAG	A	505	1	-	0/6/23/26	0/1/1/1
3	NAG	A	503	1	-	0/6/23/26	0/1/1/1
3	NAG	A	508	1	-	2/6/23/26	0/1/1/1
3	NAG	A	504	1	-	0/6/23/26	0/1/1/1
3	NAG	A	501	1	-	2/6/23/26	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	502	NAG	C1-C2	2.42	1.56	1.52
3	A	503	NAG	C1-C2	2.09	1.55	1.52

There are no bond angle outliers.

There are no chirality outliers.

All (6) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	508	NAG	O5-C5-C6-O6
3	A	508	NAG	C4-C5-C6-O6
3	A	501	NAG	C8-C7-N2-C2
3	A	501	NAG	O7-C7-N2-C2
3	A	502	NAG	C8-C7-N2-C2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms
3	A	502	NAG	O7-C7-N2-C2

There are no ring outliers.

4 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	502	NAG	1	0
3	A	505	NAG	1	0
3	A	503	NAG	1	0
3	A	504	NAG	1	0

## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	337/376 (89%)	0.28	26 (7%) 13 10	25, 36, 71, 114	0
2	R	24/28 (85%)	1.29	7 (29%) 0 0	31, 59, 82, 87	0
All	All	361/404 (89%)	0.35	33 (9%) 9 7	25, 37, 75, 114	0

All (33) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	45	TRP	6.1
1	A	492	GLU	6.0
1	A	44	VAL	5.6
1	A	324	GLY	5.0
1	A	459	GLY	4.9
2	R	13	LEU	4.8
1	A	411	GLY	4.6
1	A	356	ASN	4.6
1	A	458	GLY	3.9
1	A	412	ARG	3.8
2	R	27	VAL	3.6
2	R	11[A]	LYS	3.3
1	A	301	ASN	3.3
2	R	5	PHE	3.2
2	R	15	LEU	3.1
1	A	434[A]	MET	3.0
1	A	464	GLY	2.9
1	A	396	THR	2.9
1	A	226	LEU	2.8
1	A	232	LYS	2.6
1	A	80	ASN	2.6
2	R	16	LEU	2.5
1	A	87	GLU	2.5
1	A	395	ASP	2.4

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	463	ASN	2.3
1	A	82	GLN	2.3
1	A	355	ASN	2.2
1	A	268	GLU	2.2
1	A	88	ASN	2.2
1	A	267	GLU	2.1
1	A	49	THR	2.1
2	R	14	GLY	2.1
1	A	90[A]	THR	2.0

## 6.2 Non-standard residues in protein, DNA, RNA chains [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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
2	U3X	R	23[A]	19/20	0.97	0.11	28,36,47,48	38
2	U3X	R	23[B]	19/20	0.97	0.11	23,33,40,43	38
2	DPR	R	21	7/8	0.97	0.07	29,33,40,43	0

## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
3	NAG	A	508	14/15	0.72	0.35	57,73,87,101	0
3	NAG	A	502	14/15	0.80	0.30	59,70,84,88	0
3	NAG	A	503	14/15	0.81	0.26	57,70,76,81	0
3	NAG	A	505	14/15	0.86	0.19	54,71,81,86	0
3	NAG	A	506	14/15	0.91	0.14	46,58,64,80	0
3	NAG	A	507	14/15	0.93	0.18	48,62,73,76	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
3	NAG	A	504	14/15	0.95	0.08	27,32,48,50	0
3	NAG	A	501	14/15	0.96	0.10	32,41,50,60	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.