wwPDB X-ray Structure Validation Summary Report (i)

Sep 12, 2023 – 09:24 PM EDT

PDB ID	:	4N21
Title	:	Crystal structure of the GP2 Core Domain from the California Academy of Science Virus
Authors	:	Malashkevich, V.N.; Koellhoffer, J.F.; Dai, Z.; Toro, R.; Lai, J.R.; Almo, S.C.; New York Structural Genomics Research Consortium (NYSGRC)
Deposited on	:	2013-10-04
Resolution	:	1.99 Å(reported)

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	:	2.35.1
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 1.99 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.
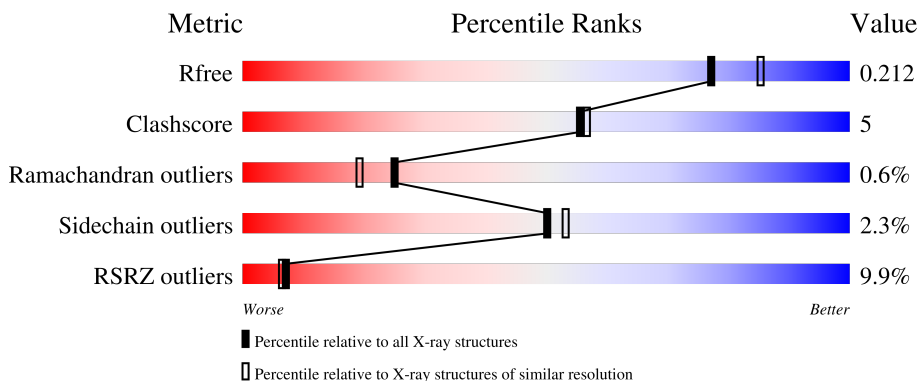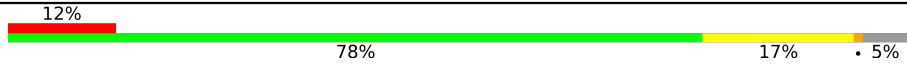

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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	130	9% / 84% 11% 5%
1	B	130	11% / 83% 11% • 5%
1	C	130	8% / 87% 8% 5%
1	D	130	6% / 77% 16% • 5%
1	E	130	10% / 85% 8% 7%

Continued on next page...

Continued from previous page...

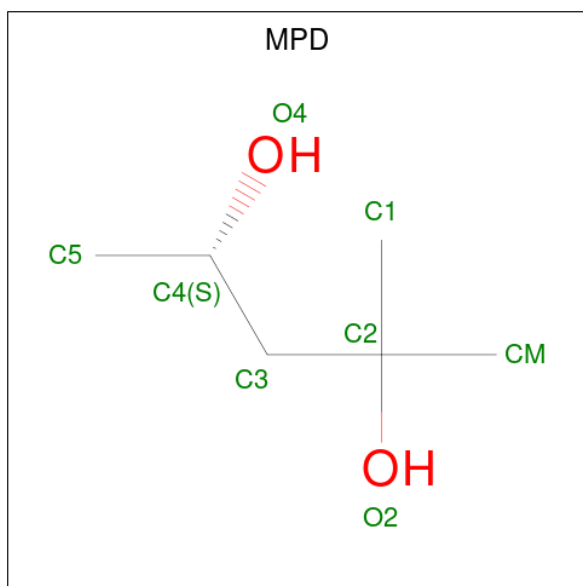| Mol	Chain	Length	Quality of chain
1	F	130	12% ▬▬▬ / 78% / 17% / 5%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 6377 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 GP2 Ectodomain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	123	Total	C	N	O	S	0	4	0
			1011	642	167	196	6			
1	B	123	Total	C	N	O	S	0	2	0
			1004	637	166	195	6			
1	C	124	Total	C	N	O	S	0	4	0
			1023	649	169	199	6			
1	D	123	Total	C	N	O	S	0	5	0
			1016	645	167	198	6			
1	E	121	Total	C	N	O	S	0	2	0
			986	627	164	189	6			
1	F	124	Total	C	N	O	S	0	3	0
			1017	645	169	197	6			

- Molecule 2 is (4S)-2-METHYL-2,4-PENTANEDIOL (three-letter code: MPD) (formula: $C_6H_{14}O_2$).

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total 8	C 6	O 2	0	0
2	A	1	Total 8	C 6	O 2	0	0
2	B	1	Total 8	C 6	O 2	0	0
2	B	1	Total 8	C 6	O 2	0	0
2	B	1	Total 8	C 6	O 2	0	0
2	C	1	Total 8	C 6	O 2	0	0
2	C	1	Total 8	C 6	O 2	0	0
2	D	1	Total 8	C 6	O 2	0	0
2	D	1	Total 8	C 6	O 2	0	0
2	E	1	Total 8	C 6	O 2	0	0
2	E	1	Total 8	C 6	O 2	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	38	Total 38	O 38	0	0
3	B	37	Total 37	O 37	0	0
3	C	37	Total 37	O 37	0	0
3	D	29	Total 29	O 29	0	0
3	E	47	Total 47	O 47	0	0
3	F	44	Total 44	O 44	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 o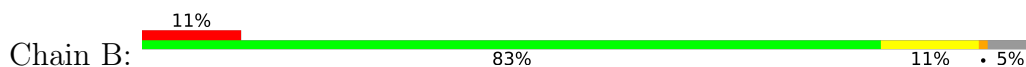r more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: GP2 Ectodomain
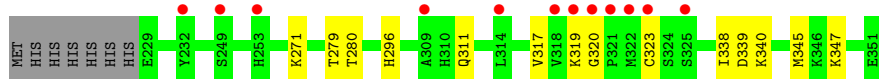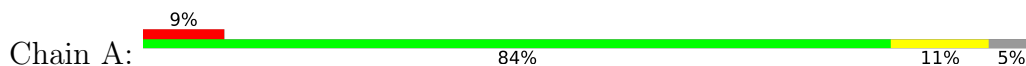
Chain A:

- Molecule 1: GP2 Ectodomain

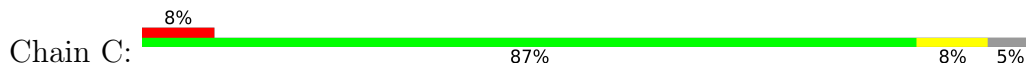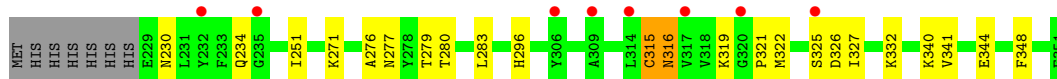Chain B:

- Molecule 1: GP2 Ectodomain

Chain C:

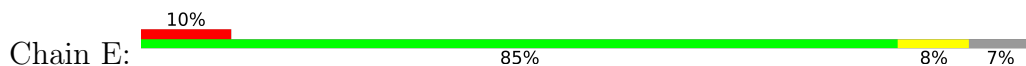- Molecule 1: GP2 Ectodomain
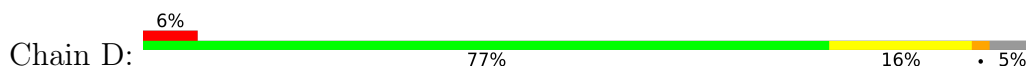
Chain D:

- Molecule 1: GP2 Ectodomain

Chain E:

- Molecule 1: GP2 Ectodomain

Chain F:

12%

78%

17%

5%

MET HIS HIS HIS HIS H228 E229 N230 L231 Y232 F233 Q234 G235 Q239 H253 A267 K271 A276 N277 T280 L283 E301 L302 Y306 A309 L314 C315 N316 V317 V318 K319 G320 P321 M322 C323 S324 S325 D326 D336 M337 K340 M345 E351

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	46.68Å 46.71Å 109.45Å 102.32° 89.99° 119.98°	Depositor
Resolution (Å)	39.22 – 1.99 39.22 – 1.99	Depositor EDS
% Data completeness (in resolution range)	97.1 (39.22-1.99) 97.1 (39.22-1.99)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.76 (at 2.00Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.168 , 0.212 0.168 , 0.212	Depositor DCC
R_{free} test set	2620 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å^2)	30.5	Xtriage
Anisotropy	0.094	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 43.0	EDS
L-test for twinning [2]	$<|L|> = 0.55, <L^2> = 0.39$	Xtriage
Estimated twinning fraction	0.457 for -h-k,h,k+l 0.457 for k,-h-k,h+k+l 0.450 for h,-h-k,-l 0.449 for k,h,-h-k-l 0.448 for -h-k,k,-k-l	Xtriage
Reported twinning fraction	0.213 for H, K, L 0.151 for h,-h-k,-l 0.161 for -H-K, K, -K-L 0.168 for K, -H-K, H+K+L 0.146 for -H-K, H, K+L 0.161 for K, H, -H-K-L	Depositor
Outliers	0 of 51571 reflections	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	6377	wwPDB-VP
Average B, all atoms (Å^2)	53.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.11% 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: MPD

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	0/1038	0.57	0/1394
1	B	0.45	0/1025	0.55	0/1376
1	C	0.45	0/1051	0.55	0/1412
1	D	0.46	0/1046	0.54	0/1405
1	E	0.45	0/1007	0.57	0/1352
1	F	0.46	0/1042	0.55	0/1399
All	All	0.46	0/6209	0.56	0/8338

There are no bond length outliers.

There are no bond angle outliers.

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	1011	0	1020	14	0
1	B	1004	0	1006	9	0
1	C	1023	0	1024	9	0
1	D	1016	0	1024	18	0
1	E	986	0	992	7	0
1	F	1017	0	1018	16	0
2	A	16	0	28	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	24	0	42	2	0
2	C	16	0	28	1	0
2	D	16	0	28	1	0
2	E	16	0	28	2	0
3	A	38	0	0	2	0
3	B	37	0	0	1	0
3	C	37	0	0	1	0
3	D	29	0	0	0	0
3	E	47	0	0	1	0
3	F	44	0	0	6	0
All	All	6377	0	6238	63	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 63 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:280[B]:THR:CG2	1:A:338:ILE:HG23	2.10	0.81
1:F:336:ASP:HB2	3:F:405:HOH:O	1.84	0.76
1:A:340:LYS:HE3	2:B:402:MPD:HM2	1.69	0.75
1:D:276:ALA:O	1:D:280[A]:THR:HG23	1.87	0.74
1:F:277:ASN:HB2	3:F:413:HOH:O	1.91	0.69

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	125/130 (96%)	122 (98%)	2 (2%)	1 (1%)	19	13
1	B	123/130 (95%)	119 (97%)	3 (2%)	1 (1%)	19	13

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	126/130 (97%)	124 (98%)	2 (2%)	0	100	100
1	D	126/130 (97%)	122 (97%)	4 (3%)	0	100	100
1	E	121/130 (93%)	116 (96%)	4 (3%)	1 (1%)	19	13
1	F	125/130 (96%)	123 (98%)	1 (1%)	1 (1%)	19	13
All	All	746/780 (96%)	726 (97%)	16 (2%)	4 (0%)	25	23

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	319	LYS
1	B	319	LYS
1	E	319	LYS
1	F	319	LYS

5.3.2 Protein sidechains ⓘ

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	115/118 (98%)	115 (100%)	0	100	100
1	B	113/118 (96%)	109 (96%)	4 (4%)	36	35
1	C	116/118 (98%)	114 (98%)	2 (2%)	60	65
1	D	116/118 (98%)	112 (97%)	4 (3%)	37	36
1	E	111/118 (94%)	108 (97%)	3 (3%)	44	46
1	F	115/118 (98%)	113 (98%)	2 (2%)	60	65
All	All	686/708 (97%)	671 (98%)	15 (2%)	50	55

5 of 15 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	316	ASN
1	F	232	TYR
1	D	332	LYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	F	302	LEU
1	E	301	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	256	ASN
1	B	350	HIS
1	D	230	ASN
1	E	342	HIS

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 ⓘ

11 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	$\#|Z| > 2$
2	MPD	A	402	-	7,7,7	0.14	0	9,10,10	0.46	0
2	MPD	C	402	-	7,7,7	0.29	0	9,10,10	0.44	0
2	MPD	D	401	-	7,7,7	0.21	0	9,10,10	0.52	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	MPD	B	403	-	7,7,7	0.32	0	9,10,10	0.45	0
2	MPD	B	401	-	7,7,7	0.23	0	9,10,10	0.61	0
2	MPD	D	402	-	7,7,7	0.31	0	9,10,10	0.31	0
2	MPD	A	401	-	7,7,7	0.27	0	9,10,10	0.59	0
2	MPD	C	401	-	7,7,7	0.29	0	9,10,10	0.46	0
2	MPD	E	401	-	7,7,7	0.46	0	9,10,10	0.59	0
2	MPD	E	402	-	7,7,7	0.20	0	9,10,10	0.34	0
2	MPD	B	402	-	7,7,7	0.25	0	9,10,10	0.62	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	MPD	A	402	-	-	2/5/5/5	-
2	MPD	C	402	-	-	1/5/5/5	-
2	MPD	D	401	-	-	2/5/5/5	-
2	MPD	B	403	-	-	1/5/5/5	-
2	MPD	B	401	-	-	2/5/5/5	-
2	MPD	D	402	-	-	3/5/5/5	-
2	MPD	A	401	-	-	2/5/5/5	-
2	MPD	C	401	-	-	1/5/5/5	-
2	MPD	E	401	-	-	3/5/5/5	-
2	MPD	E	402	-	-	1/5/5/5	-
2	MPD	B	402	-	-	2/5/5/5	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 20 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	D	402	MPD	C2-C3-C4-O4
2	E	401	MPD	C2-C3-C4-C5
2	B	401	MPD	O2-C2-C3-C4
2	D	402	MPD	O2-C2-C3-C4
2	B	403	MPD	C2-C3-C4-C5

There are no ring outliers.

6 monomers are involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	402	MPD	1	0
2	C	402	MPD	1	0
2	D	401	MPD	1	0
2	A	401	MPD	1	0
2	E	401	MPD	2	0
2	B	402	MPD	2	0

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	123/130 (94%)	0.58	12 (9%)	7	7	25, 48, 85, 112	0
1	B	123/130 (94%)	0.65	14 (11%)	5	4	25, 49, 101, 121	0
1	C	124/130 (95%)	0.54	11 (8%)	9	8	27, 49, 96, 114	0
1	D	123/130 (94%)	0.52	8 (6%)	18	18	27, 48, 84, 95	0
1	E	121/130 (93%)	0.78	13 (10%)	6	5	25, 47, 105, 114	0
1	F	124/130 (95%)	0.50	15 (12%)	4	3	25, 50, 95, 111	0
All	All	738/780 (94%)	0.60	73 (9%)	7	6	25, 49, 97, 121	0

The worst 5 of 73 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	317	VAL	9.0
1	B	320	GLY	8.4
1	C	320	GLY	7.8
1	D	325	SER	7.7
1	E	313	GLY	7.3

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.

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
2	MPD	C	401	8/8	0.89	0.20	57,67,70,72	0
2	MPD	E	402	8/8	0.89	0.16	48,49,52,58	0
2	MPD	C	402	8/8	0.91	0.14	50,53,58,60	0
2	MPD	A	401	8/8	0.92	0.19	50,54,55,58	0
2	MPD	B	403	8/8	0.93	0.16	58,60,63,67	0
2	MPD	A	402	8/8	0.93	0.13	48,54,56,61	0
2	MPD	E	401	8/8	0.94	0.17	44,46,48,49	0
2	MPD	B	402	8/8	0.94	0.14	48,50,56,68	0
2	MPD	D	401	8/8	0.95	0.17	49,54,55,59	0
2	MPD	B	401	8/8	0.96	0.14	46,52,56,60	0
2	MPD	D	402	8/8	0.96	0.13	49,58,59,61	0

6.5 Other polymers ⓘ

There are no such residues in this entry.