WORLDWIDE PDB

PROTEIN DATA BANK

Full wwPDB X-ray Structure Validation Report ⓘ

Sep 13, 2023 – 01:31 AM EDT

PDB ID	:	4N23
Title	:	Crystal structure of the GP2 Core Domain from the California Academy of Science Virus, monoclinic symmetry
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	:	2.00 Å(reported)

The following versions of software and data (see references ⓘ) were used in the production of this report:

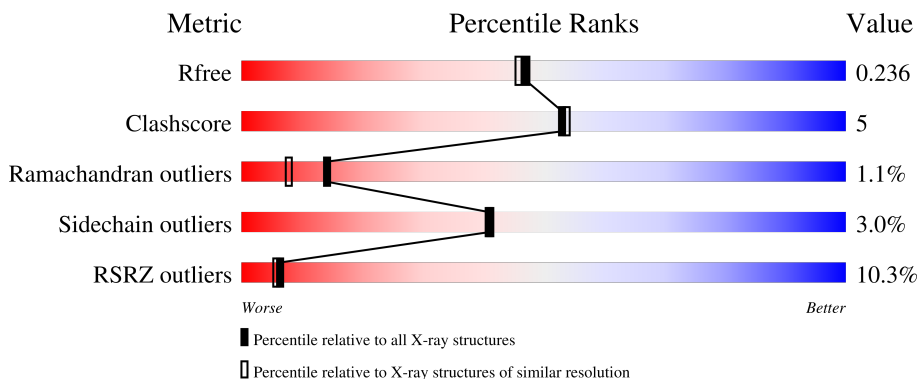|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (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 2.00 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		0.236
Clashscore		5
Ramachandran outliers		1.1%
Sidechain outliers		3.0%
RSRZ outliers		10.3%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

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.

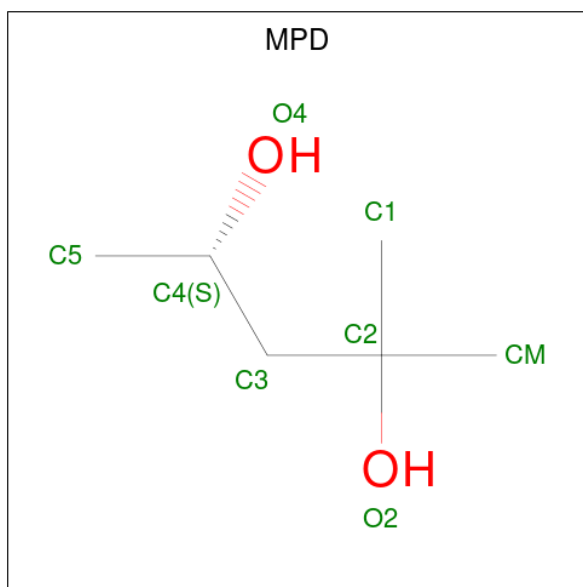| Mol	Chain	Length	Quality of chain
1	A	130	7% 76% · 18% · 5%
1	B	130	12% 85% · 8% · 6%
1	C	130	10% 81% · 15% · 5%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 3188 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	2	0
			1003	636	167	194	6			
1	B	122	Total	C	N	O	S	0	3	0
			1002	637	167	192	6			
1	C	124	Total	C	N	O	S	0	1	0
			1012	642	171	193	6			

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

MPD

O4
OH
C1
C5
C4(S)
C2
CM
C3
OH
O2
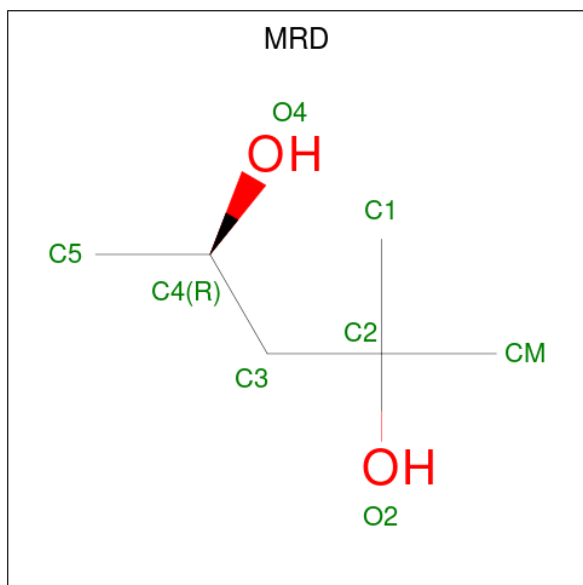
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			8	6	2		
2	B	1	Total	C	O	0	0
			8	6	2		
2	C	1	Total	C	O	0	0
			8	6	2		

Continued on next page...

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

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	C	1	Total 8	C 6	O 2	0	0

- Molecule 3 is (4R)-2-METHYLPENTANE-2,4-DIOL (three-letter code: MRD) (formula: $C_6H_{14}O_2$).

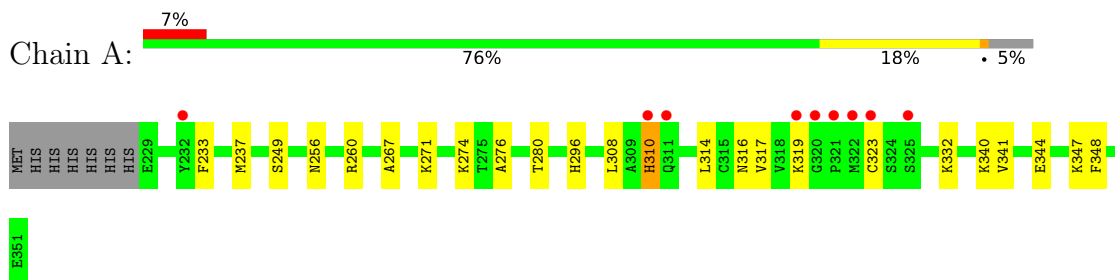| Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	C	1	Total 8	C 6	O 2	0	0

- Molecule 4 is water.

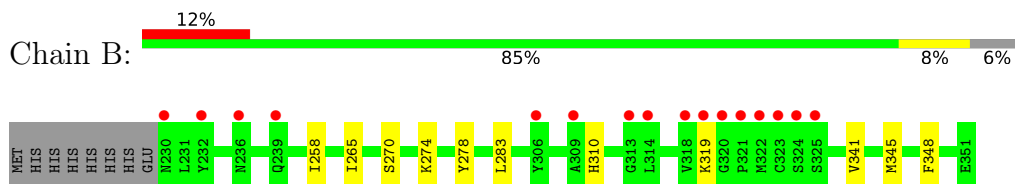| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	40	Total 40	O 40	0	0
4	B	42	Total 42	O 42	0	0
4	C	49	Total 49	O 49	0	0

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.
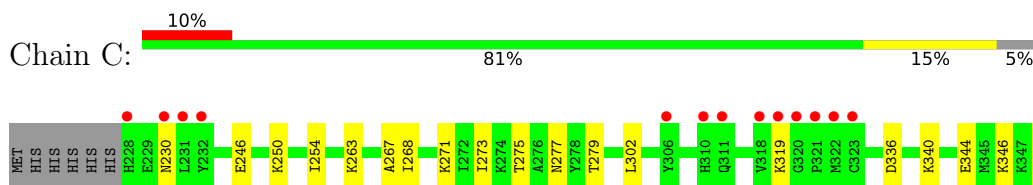
- Molecule 1: GP2 Ectodomain

Chain A:

- Molecule 1: GP2 Ectodomain

Chain B:

- Molecule 1: GP2 Ectodomain

Chain C:

4 Data and refinement statistics   ⓘ

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	80.87Å 46.72Å 109.47Å 90.00° 104.28° 90.00°	Depositor
Resolution (Å)	39.21 – 2.00 39.21 – 2.00	Depositor EDS
% Data completeness (in resolution range)	99.6 (39.21-2.00) 99.6 (39.21-2.00)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.33 (at 2.00Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.176 , 0.228 0.177 , 0.236	Depositor DCC
R_{free} test set	1352 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	30.7	Xtriage
Anisotropy	0.133	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 33.8	EDS
L-test for twinning[2]	$< |L| > = 0.53, < L^2 > = 0.37$	Xtriage
Estimated twinning fraction	0.469 for 1/2*h+3/2*k,1/2*h-1/2*k,-1/2*h-1/2*k-l 0.468 for 1/2*h-3/2*k,-1/2*h-1/2*k,-1/2*h+1/2*k-l	Xtriage
Reported twinning fraction	0.299 for H, K, L 0.347 for 1/2H+3/2K, 1/2H-1/2K, -1/2H-1/2K-L 0.354 for 1/2H-3/2K, -1/2H-1/2K, -1/2H+1/2K-L	Depositor
Outliers	2 of 27014 reflections (0.007%)	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	3188	wwPDB-VP
Average B, all atoms (Å^2)	52.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.68% 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: MRD, 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	0.45	0/1024	0.57	0/1374
1	B	0.45	0/1026	0.54	0/1376
1	C	0.46	0/1032	0.54	0/1385
All	All	0.45	0/3082	0.55	0/4135

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	1003	0	1006	14	0
1	B	1002	0	1010	8	0
1	C	1012	0	1009	11	0
2	A	8	0	14	0	0
2	B	8	0	14	3	0
2	C	16	0	28	2	0
3	C	8	0	14	3	0
4	A	40	0	0	2	0
4	B	42	0	0	1	0
4	C	49	0	0	3	0
All	All	3188	0	3095	33	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.

All (33) 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:C:402:MPD:O4	2:C:402:MPD:H12	1.66	0.94
1:A:280:THR:HG21	1:A:341:VAL:HG11	1.61	0.82
3:C:403:MRD:H1C2	3:C:403:MRD:O4	1.80	0.80
1:A:314:LEU:HA	1:A:317:VAL:HG12	1.68	0.75
1:B:274[A]:LYS:HE2	2:B:401:MPD:H51	1.67	0.75
3:C:403:MRD:H5C1	4:C:526:HOH:O	1.89	0.73
1:C:246:GLU:O	1:C:250:LYS:HG2	1.91	0.69
1:B:345:MET:HG3	4:B:514:HOH:O	1.95	0.65
1:A:280:THR:CG2	1:A:341:VAL:HG11	2.27	0.63
1:A:280:THR:HG21	1:A:341:VAL:CG1	2.31	0.60
1:A:344:GLU:O	1:A:347:LYS:HB2	2.03	0.59
1:A:276:ALA:O	1:A:280:THR:HG23	2.06	0.55
1:C:267:ALA:O	1:C:271:LYS:HG3	2.08	0.54
1:A:308:LEU:HG	1:A:317:VAL:HG11	1.89	0.52
1:A:310:HIS:HA	4:A:537:HOH:O	2.11	0.51
1:C:263:LYS:NZ	4:C:543:HOH:O	2.39	0.50
1:B:345:MET:CE	1:B:348:PHE:HE2	2.26	0.49
1:C:277:ASN:HB2	4:C:527:HOH:O	2.14	0.47
1:B:283:LEU:HD11	1:C:279:THR:HG23	1.98	0.46
1:A:233:PHE:O	1:A:237:MET:HB2	2.15	0.46
1:B:258:ILE:HD11	1:C:254:ILE:HG23	1.99	0.45
1:A:256:ASN:O	1:A:260:ARG:HG3	2.17	0.45
1:A:296:HIS:HE1	4:A:538:HOH:O	2.00	0.45
3:C:403:MRD:O4	3:C:403:MRD:C1	2.54	0.44
1:B:278:TYR:OH	2:B:401:MPD:HM1	2.18	0.43
1:A:274:LYS:HE3	1:C:344:GLU:CD	2.39	0.43
1:C:336:ASP:OD1	1:C:340:LYS:HE2	2.18	0.43
2:C:402:MPD:H12	2:C:402:MPD:HO4	1.79	0.42
1:C:273:ILE:HD11	1:C:348:PHE:CB	2.50	0.41
1:B:341:VAL:HG13	1:C:275:THR:HG23	2.02	0.41
1:A:340:LYS:HD3	2:B:401:MPD:CM	2.51	0.40
1:B:265:ILE:HG23	1:C:268:ILE:HD11	2.03	0.40
1:A:267:ALA:O	1:A:271:LYS:HG2	2.22	0.40

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	123/130 (95%)	115 (94%)	6 (5%)	2 (2%)	9	4
1	B	123/130 (95%)	120 (98%)	2 (2%)	1 (1%)	19	13
1	C	123/130 (95%)	118 (96%)	4 (3%)	1 (1%)	19	13
All	All	369/390 (95%)	353 (96%)	12 (3%)	4 (1%)	14	8

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	319	LYS
1	A	323	CYS
1	B	319	LYS
1	C	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	113/118 (96%)	108 (96%)	5 (4%)	28	25
1	B	113/118 (96%)	111 (98%)	2 (2%)	59	63
1	C	113/118 (96%)	110 (97%)	3 (3%)	44	46
All	All	339/354 (96%)	329 (97%)	10 (3%)	41	43

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

Mol	Chain	Res	Type
1	A	249	SER
1	A	310	HIS
1	A	316	ASN
1	A	332	LYS
1	A	348	PHE
1	B	270	SER
1	B	310	HIS
1	C	230	ASN
1	C	302	LEU
1	C	346	LYS

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

Mol	Chain	Res	Type
1	A	311	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 ⓘ

5 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	401	-	7,7,7	0.32	0	9,10,10	0.51	0
2	MPD	C	402	-	7,7,7	0.19	0	9,10,10	0.25	0
2	MPD	C	401	-	7,7,7	0.29	0	9,10,10	0.60	0
2	MPD	B	401	-	7,7,7	0.27	0	9,10,10	0.51	0
3	MRD	C	403	-	7,7,7	0.31	0	9,10,10	0.30	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	401	-	-	3/5/5/5	-
2	MPD	C	402	-	-	1/5/5/5	-
2	MPD	C	401	-	-	2/5/5/5	-
2	MPD	B	401	-	-	1/5/5/5	-
3	MRD	C	403	-	-	2/5/5/5	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (9) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	401	MPD	O2-C2-C3-C4
2	A	401	MPD	C2-C3-C4-C5
3	C	403	MRD	C2-C3-C4-C5
3	C	403	MRD	C2-C3-C4-O4
2	B	401	MPD	C2-C3-C4-C5
2	C	401	MPD	C2-C3-C4-C5
2	A	401	MPD	C2-C3-C4-O4
2	C	401	MPD	C2-C3-C4-O4
2	C	402	MPD	C2-C3-C4-O4

There are no ring outliers.

3 monomers are involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	C	402	MPD	2	0
2	B	401	MPD	3	0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	C	403	MRD	3	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(Å^2)	Q<0.9
1	A	123/130 (94%)	0.49	9 (7%)	15	14	21, 44, 98, 144	0
1	B	122/130 (93%)	0.48	16 (13%)	3	3	19, 46, 99, 124	0
1	C	124/130 (95%)	0.33	13 (10%)	6	5	20, 47, 105, 129	0
All	All	369/390 (94%)	0.43	38 (10%)	6	6	19, 45, 104, 144	0

All (38) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	322	MET	11.2
1	A	325	SER	9.5
1	A	321	PRO	8.0
1	C	321	PRO	6.3
1	C	318	VAL	6.2
1	C	320	GLY	5.9
1	C	319	LYS	5.7
1	B	232	TYR	5.7
1	A	320	GLY	5.6
1	B	309	ALA	5.0
1	B	322	MET	5.0
1	B	319	LYS	4.9
1	C	232	TYR	4.6
1	B	321	PRO	4.5
1	B	320	GLY	4.4
1	C	322	MET	4.3
1	B	324	SER	4.2
1	B	306	TYR	4.1
1	A	232	TYR	4.0
1	C	310[A]	HIS	3.7
1	B	325	SER	3.6
1	A	319	LYS	3.3
1	B	314	LEU	3.0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	313	GLY	2.8
1	A	311	GLN	2.7
1	C	311	GLN	2.7
1	C	323	CYS	2.7
1	B	323	CYS	2.6
1	B	318	VAL	2.5
1	C	228	HIS	2.5
1	A	310	HIS	2.4
1	C	230	ASN	2.3
1	C	306	TYR	2.2
1	A	323	CYS	2.2
1	B	230	ASN	2.2
1	B	239	GLN	2.1
1	C	231	LEU	2.1
1	B	236	ASN	2.0

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($\mathring{A}^2$)	Q<0.9
2	MPD	B	401	8/8	0.84	0.16	48,51,54,57	0
2	MPD	C	401	8/8	0.89	0.15	41,49,53,56	0
2	MPD	A	401	8/8	0.93	0.17	44,46,48,50	0
2	MPD	C	402	8/8	0.94	0.14	41,43,44,45	8
3	MRD	C	403	8/8	0.95	0.13	33,36,36,38	8

6.5 Other polymers ⓘ

There are no such residues in this entry.