



wwPDB X-ray Structure Validation Summary Report ⓘ

May 31, 2022 – 06:11 pm BST

PDB ID : 7PB8
Title : Crystal structure of the CENP-OPQUR complex
Authors : Bellini, D.; Yatskevich, S.; Muir, K.W.; Barford, D.
Deposited on : 2021-07-31
Resolution : 3.68 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

<https://www.wwpdb.org/validation/2017/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
Xtriage (Phenix) : 1.13
EDS : 2.28.1
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac : 5.8.0267
CCP4 : 7.1.010 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.28.1

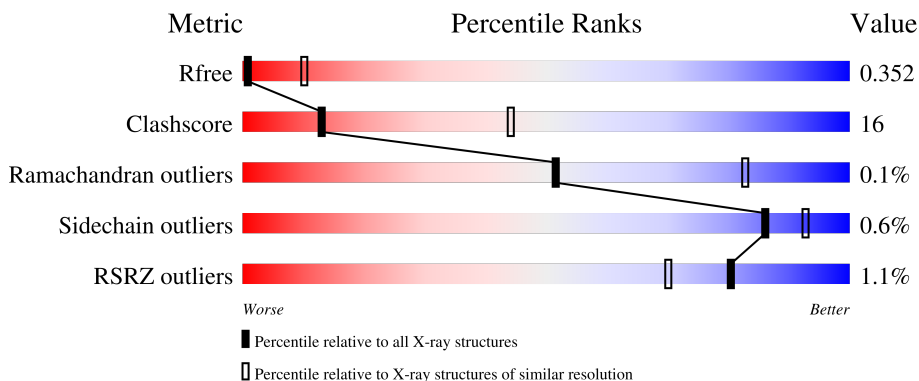
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.68 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	1013 (3.84-3.52)
Clashscore	141614	1070 (3.84-3.52)
Ramachandran outliers	138981	1036 (3.84-3.52)
Sidechain outliers	138945	1033 (3.84-3.52)
RSRZ outliers	127900	1471 (3.86-3.50)

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 $\leq 5\%$. 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	O	300	 45% 20% 34%
2	Q	137	 55% 27% 18%
3	U	418	 17% 7% 75%
4	P	294	 44% 32% 24%
5	R	177	 24% 15% 61%

2 Entry composition i

There are 6 unique types of molecules in this entry. The entry contains 5706 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 Centromere protein O.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	O	198	1586	1020	265	294	7	0	0	0

- Molecule 2 is a protein called Centromere protein Q.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	Q	112	886	555	145	181	5	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
Q	132	MET	-	initiating methionine	UNP Q7L2Z9

- Molecule 3 is a protein called Centromere protein U.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	U	103	858	538	154	165	1	0	0	0

- Molecule 4 is a protein called Centromere protein P.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	P	224	1818	1158	316	336	8	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
P	289	GLU	-	expression tag	UNP Q6IPU0
P	290	ASN	-	expression tag	UNP Q6IPU0
P	291	LEU	-	expression tag	UNP Q6IPU0

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
P	292	TYR	-	expression tag	UNP Q6IPU0
P	293	PHE	-	expression tag	UNP Q6IPU0
P	294	GLN	-	expression tag	UNP Q6IPU0

- Molecule 5 is a protein called Centromere protein R.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	R	69	557	351	92	107	7	0	0	0

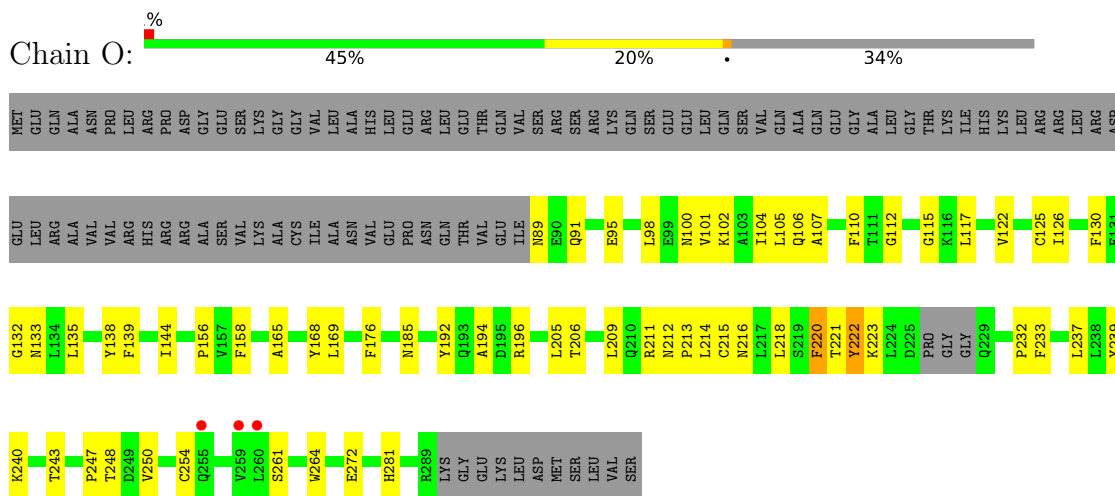
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
6	U	1	1	1	0	0

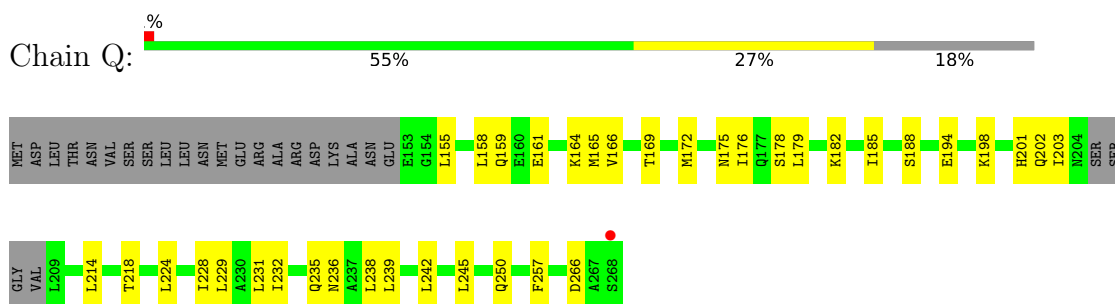
3 Residue-property plots [i](#)

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.

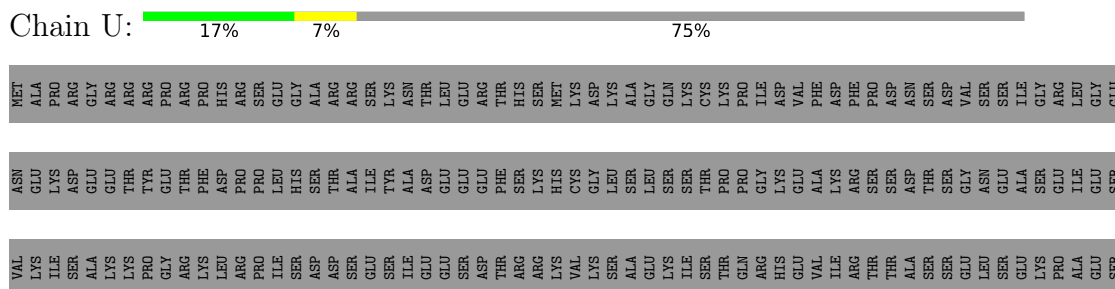
- Molecule 1: Centromere protein O



- Molecule 2: Centromere protein Q



- Molecule 3: Centromere protein U



4 Data and refinement statistics i

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	199.61Å 51.30Å 110.00Å 90.00° 117.49° 90.00°	Depositor
Resolution (Å)	25.65 – 3.68 25.65 – 3.68	Depositor EDS
% Data completeness (in resolution range)	55.4 (25.65-3.68) 55.4 (25.65-3.68)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.17 (at 3.64Å)	Xtrriage
Refinement program	REFMAC 1.17.1_3660, PHENIX 1.17.1_3660	Depositor
R, R_{free}	0.256 , 0.352 0.257 , 0.352	Depositor DCC
R_{free} test set	295 reflections (4.83%)	wwPDB-VP
Wilson B-factor (Å ²)	79.9	Xtrriage
Anisotropy	0.133	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	(Not available) , (Not available)	EDS
L-test for twinning ²	$\langle L \rangle = 0.45$, $\langle L^2 \rangle = 0.27$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.86	EDS
Total number of atoms	5706	wwPDB-VP
Average B, all atoms (Å ²)	100.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.52% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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](#)

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	O	0.33	0/1622	0.52	0/2202
2	Q	0.27	0/891	0.45	0/1195
3	U	0.27	0/869	0.43	0/1166
4	P	0.32	0/1850	0.50	0/2487
5	R	0.30	0/560	0.57	0/741
All	All	0.31	0/5792	0.50	0/7791

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	O	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	O	221	THR	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	O	1586	0	1572	55	0
2	Q	886	0	918	29	0
3	U	858	0	875	36	0
4	P	1818	0	1845	84	0
5	R	557	0	584	29	0
6	U	1	0	0	0	0
All	All	5706	0	5794	188	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 16.

The worst 5 of 188 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)
4:P:86:GLU:OE2	4:P:104:ARG:NH1	2.11	0.83
3:U:412:GLU:HA	3:U:415:LEU:HB2	1.61	0.83
1:O:223:LYS:HG2	1:O:232:PRO:HA	1.64	0.79
4:P:217:GLU:HG2	4:P:241:LYS:HD3	1.64	0.77
2:Q:161:GLU:HG2	2:Q:164:LYS:HD3	1.68	0.75

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	O	194/300 (65%)	173 (89%)	21 (11%)	0	100	100
2	Q	108/137 (79%)	103 (95%)	4 (4%)	1 (1%)	17	54
3	U	101/418 (24%)	97 (96%)	4 (4%)	0	100	100
4	P	220/294 (75%)	204 (93%)	16 (7%)	0	100	100
5	R	67/177 (38%)	63 (94%)	4 (6%)	0	100	100
All	All	690/1326 (52%)	640 (93%)	49 (7%)	1 (0%)	51	83

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	Q	203	ILE

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	O	177/263 (67%)	175 (99%)	2 (1%)	73	85
2	Q	103/125 (82%)	103 (100%)	0	100	100
3	U	96/379 (25%)	96 (100%)	0	100	100
4	P	205/265 (77%)	203 (99%)	2 (1%)	76	86
5	R	65/166 (39%)	65 (100%)	0	100	100
All	All	646/1198 (54%)	642 (99%)	4 (1%)	86	93

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	O	220	PHE
1	O	222	TYR
4	P	116	PHE
4	P	263	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
4	P	112	HIS
5	R	147	ASN
3	U	360	ASN
3	U	408	ASN
4	P	77	ASN

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 monosaccharides in this entry.

5.6 Ligand geometry [i](#)

There are no ligands in this entry.

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, 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	O	198/300 (66%)	-0.38	3 (1%) 73 62	43, 73, 138, 174	0
2	Q	112/137 (81%)	-0.34	1 (0%) 84 75	57, 115, 170, 183	0
3	U	103/418 (24%)	-0.26	0 100 100	55, 104, 166, 173	0
4	P	224/294 (76%)	-0.31	3 (1%) 77 66	49, 86, 131, 167	0
5	R	69/177 (38%)	-0.04	1 (1%) 75 63	137, 160, 174, 179	0
All	All	706/1326 (53%)	-0.30	8 (1%) 80 70	43, 93, 167, 183	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	P	144	PRO	4.9
1	O	260	LEU	3.8
1	O	259	VAL	3.2
5	R	115	GLY	3.0
4	P	145	THR	2.8

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 monosaccharides in this entry.

6.4 Ligands [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.