WORLDWIDE PDB
PROTEIN DATA BANK

Full wwPDB X-ray Structure Validation Report (i)

Jun 28, 2022 – 10:05 pm BST

PDB ID	:	7PJO
Title	:	Crystal form 3 of CPR-C4: a cysteine protease from the Candidate Phyla Radiation
Authors	:	Cornish, K.A.S.; Pohl, E.
Deposited on	:	2021-08-24
Resolution	:	2.25 Å(reported)

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

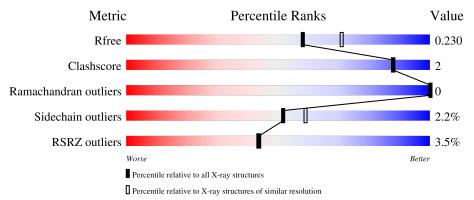|		
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.29
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.29

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 2.25 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		0.230
Clashscore		2
Ramachandran outliers		0
Sidechain outliers		2.2%
RSRZ outliers		3.5%

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	2391 (2.26-2.22)
Clashscore	141614	2539 (2.26-2.22)
Ramachandran outliers	138981	2489 (2.26-2.22)
Sidechain outliers	138945	2490 (2.26-2.22)
RSRZ outliers	127900	2353 (2.26-2.22)

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.

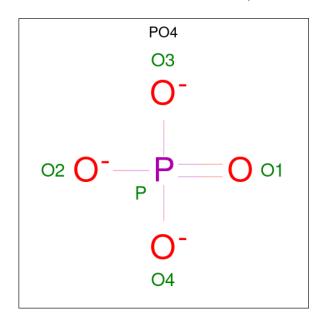| Mol	Chain	Length	Quality of chain
1	AAA	237	3% — 85% — 6% — 9%
1	BBB	237	3% — 85% — 7% — 8%

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 7215 atoms, of which 3512 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 CPR-C4.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	AAA	216	Total 3504	C 1121	H 1757	N 308	O 309	S 9	61	0	0
1	BBB	217	Total 3510	C 1125	H 1755	N 311	O 310	S 9	66	1	0

- Molecule 2 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	AAA	1	Total 5	O 4	P 1	0	0
2	AAA	1	Total 5	O 4	P 1	0	0
2	AAA	1	Total 5	O 4	P 1	0	0
2	BBB	1	Total 5	O 4	P 1	0	0

• Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	AAA	85	Total 85	O 85	0	0
3	BBB	96	Total 96	O 96	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: CPR-C4

Chain AAA:

- Molecule 1: CPR-C4

Chain BBB:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	44.78Å 77.55Å 161.72Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	44.26 – 2.25 44.26 – 2.25	Depositor EDS
% Data completeness (in resolution range)	97.9 (44.26-2.25) 97.8 (44.26-2.25)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.21 (at 2.24Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.179 , 0.225 0.182 , 0.230	Depositor DCC
R_{free} test set	1323 reflections (4.89%)	wwPDB-VP
Wilson B-factor (Å^2)	55.9	Xtriage
Anisotropy	0.034	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	7215	wwPDB-VP
Average B, all atoms (Å^2)	61.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.46% 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 ⓘ

5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: PO4

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	AAA	0.66	0/1794	0.80	1/2425 (0.0%)
1	BBB	0.67	0/1803	0.80	0/2440
All	All	0.66	0/3597	0.80	1/4865 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	AAA	151	ARG	NE-CZ-NH2	5.14	122.87	120.30

There are no chirality outliers.

There are no planarity outliers.

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	AAA	1747	1757	1737	8	0
1	BBB	1755	1755	1727	8	0
2	AAA	15	0	0	0	0
2	BBB	5	0	0	0	0
3	AAA	85	0	0	3	0
3	BBB	96	0	0	1	0
All	All	3703	3512	3464	14	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 2.

All (14) 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:AAA:93:HIS:ND1	3:AAA:401:HOH:O	2.25	0.69
1:AAA:77:HIS:ND1	3:AAA:402:HOH:O	2.32	0.59
1:AAA:121:ILE:HG22	1:BBB:194:MET:HE2	1.86	0.58
1:AAA:48:TRP:HB3	1:AAA:52:ALA:HB3	1.92	0.51
1:AAA:194:MET:HE3	1:BBB:121:ILE:HA	1.92	0.50
1:AAA:161:ARG:NH2	3:AAA:404:HOH:O	2.44	0.49
1:BBB:48:TRP:HB3	1:BBB:52:ALA:HB3	1.93	0.49
1:BBB:77:HIS:HD2	3:BBB:448:HOH:O	2.00	0.45
1:AAA:170:ALA:CB	1:AAA:174:LEU:HD21	2.48	0.43
1:BBB:170:ALA:CB	1:BBB:174:LEU:HD21	2.48	0.42
1:BBB:204:ILE:HD12	1:BBB:204:ILE:HA	1.92	0.41
1:AAA:73:ALA:HA	1:AAA:77:HIS:O	2.20	0.41
1:BBB:-4:MET:HG3	1:BBB:1:MET:HG3	2.02	0.41
1:BBB:73:ALA:HA	1:BBB:77:HIS:O	2.20	0.41

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	AAA	212/237 (90%)	203 (96%)	9 (4%)	0	100	100
1	BBB	214/237 (90%)	204 (95%)	10 (5%)	0	100	100
All	All	426/474 (90%)	407 (96%)	19 (4%)	0	100	100

There are no Ramachandran outliers to report.

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	AAA	183/203 (90%)	180 (98%)	3 (2%)	62	70
1	BBB	182/203 (90%)	177 (97%)	5 (3%)	44	51
All	All	365/406 (90%)	357 (98%)	8 (2%)	52	59

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	AAA	13	THR
1	AAA	37	ASN
1	AAA	136	TYR
1	BBB	-2	ARG
1	BBB	37	ASN
1	BBB	58	HIS
1	BBB	88	ASP
1	BBB	136	TYR

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

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 (i)

4 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	PO4	AAA	301	-	4,4,4	1.02	0	6,6,6	0.52	0
2	PO4	AAA	303	-	4,4,4	0.57	0	6,6,6	0.56	0
2	PO4	AAA	302	-	4,4,4	0.65	0	6,6,6	0.45	0
2	PO4	BBB	301	-	4,4,4	0.55	0	6,6,6	0.49	0

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 (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 ⓘ

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	AAA	216/237 (91%)	0.43	8 (3%)	41	41	42, 57, 89, 112	0
1	BBB	217/237 (91%)	0.35	7 (3%)	47	47	41, 57, 87, 117	0
All	All	433/474 (91%)	0.39	15 (3%)	44	43	41, 57, 87, 117	0

All (15) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	AAA	13	THR	5.0
1	AAA	74	TYR	3.6
1	AAA	171	THR	3.6
1	AAA	14	GLU	3.5
1	BBB	46	THR	3.4
1	AAA	12	THR	2.9
1	BBB	17	LYS	2.8
1	BBB	171	THR	2.5
1	BBB	58	HIS	2.3
1	AAA	170	ALA	2.1
1	AAA	15	GLU	2.1
1	AAA	173	ASP	2.1
1	BBB	169	ILE	2.1
1	BBB	-5	GLN	2.0
1	BBB	12	THR	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.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

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	PO4	AAA	302	5/5	0.81	0.18	97,99,114,123	0
2	PO4	AAA	303	5/5	0.88	0.17	96,97,110,113	0
2	PO4	BBB	301	5/5	0.93	0.20	95,107,110,114	0
2	PO4	AAA	301	5/5	0.96	0.10	65,72,78,89	0

6.5 Other polymers ⓘ

There are no such residues in this entry.