wwPDB X-ray Structure Validation Summary Report (i)

Sep 21, 2021 – 05:00 am BST

PDB ID	:	6YXJ
Title	:	Crystal structure of SARS-CoV macrodomain II in complex with human Paip1
Authors	:	Lei, J.; Hilgenfeld, R.
Deposited on	:	2020-05-02
Resolution	:	3.50 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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

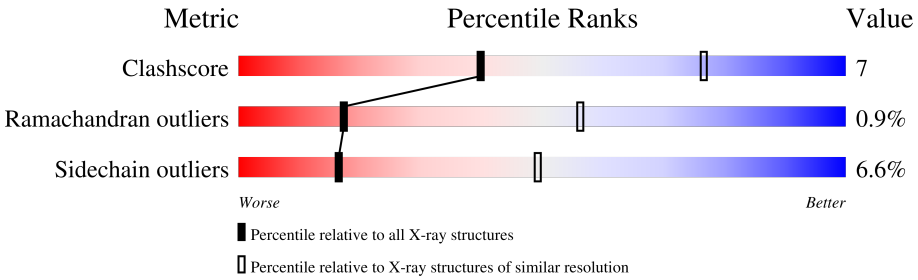
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 3.50 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))
Clashscore	141614	1036 (3.58-3.42)
Ramachandran outliers	138981	1005 (3.58-3.42)
Sidechain outliers	138945	1006 (3.58-3.42)

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%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	142	68% 18% 5% 10%
2	B	236	81% 10% • 8%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 2726 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 Non-structural protein 3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	128	Total	C	N	O	S	0	0	0
			982	626	155	193	8			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	385	GLY	-	expression tag	UNP P0C6U8
A	386	SER	-	expression tag	UNP P0C6U8
A	387	HIS	-	expression tag	UNP P0C6U8
A	388	MET	-	expression tag	UNP P0C6U8

- Molecule 2 is a protein called Polyadenylate-binding protein-interacting protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	218	Total	C	N	O	S	0	0	0
			1744	1099	297	338	10			

There are 17 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	61	GLY	-	expression tag	UNP Q9H074
B	62	SER	-	expression tag	UNP Q9H074
B	63	HIS	-	expression tag	UNP Q9H074
B	64	MET	-	expression tag	UNP Q9H074
B	65	ALA	-	expression tag	UNP Q9H074
B	66	SER	-	expression tag	UNP Q9H074
B	67	MET	-	expression tag	UNP Q9H074
B	68	THR	-	expression tag	UNP Q9H074
B	69	GLY	-	expression tag	UNP Q9H074
B	70	GLY	-	expression tag	UNP Q9H074
B	71	GLN	-	expression tag	UNP Q9H074
B	72	GLN	-	expression tag	UNP Q9H074

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	73	MET	-	expression tag	UNP Q9H074
B	74	GLY	-	expression tag	UNP Q9H074
B	75	ARG	-	expression tag	UNP Q9H074
B	76	GLY	-	expression tag	UNP Q9H074
B	77	SER	-	expression tag	UNP Q9H074

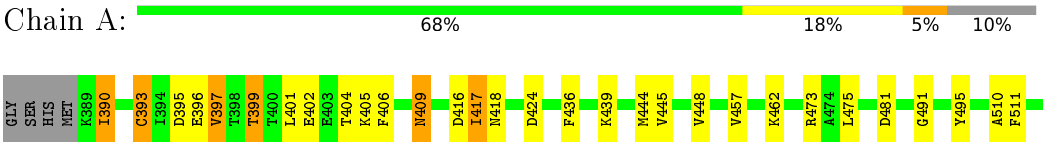
3 Residue-property plots (i)

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. 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. 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.

Note EDS was not executed.

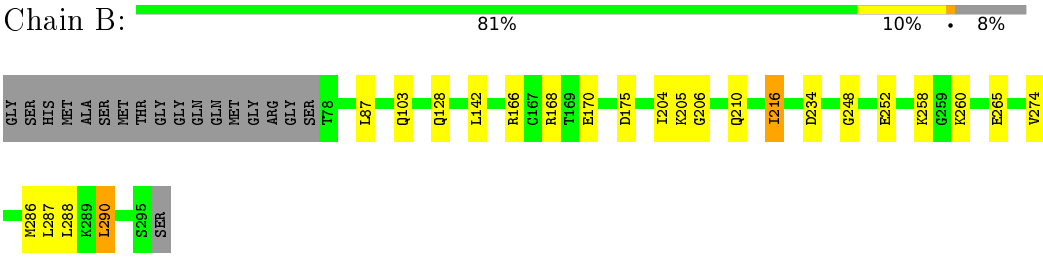
- Molecule 1: Non-structural protein 3

Chain A:



- Molecule 2: Polyadenylate-binding protein-interacting protein 1

Chain B:



4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α, β, γ	92.39Å 92.39Å 166.63Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	30.94 – 3.50	Depositor
% Data completeness (in resolution range)	98.3 (30.94-3.50)	Depositor
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	BUSTER 2.10.3 (20-MAY-2020)	Depositor
R, R_{free}	0.313 , 0.336	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	2726	wwPDB-VP
Average B, all atoms (Å^2)	255.0	wwPDB-VP

5 Model quality

5.1 Standard geometry

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	0.46	0/997	0.67	0/1345
2	B	0.34	0/1768	0.51	0/2387
All	All	0.39	0/2765	0.57	0/3732

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

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	982	0	1005	28	0
2	B	1744	0	1746	15	0
All	All	2726	0	2751	40	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 7.

The worst 5 of 40 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:436:PHE:HA	1:A:439:LYS:HZ3	1.31	0.92
1:A:406:PHE:HA	2:B:168:ARG:HD3	1.60	0.84

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:395:ASP:HA	1:A:511:PHE:O	1.86	0.75
1:A:436:PHE:HB2	1:A:439:LYS:HZ1	1.57	0.70
2:B:206:GLY:H	2:B:210:GLN:HA	1.55	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	126/142 (89%)	116 (92%)	7 (6%)	3 (2%)	6 35
2	B	216/236 (92%)	213 (99%)	3 (1%)	0	100 100
All	All	342/378 (90%)	329 (96%)	10 (3%)	3 (1%)	17 56

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	491	GLY
1	A	393	CYS
1	A	397	VAL

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.

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	110/121 (91%)	99 (90%)	11 (10%)	7 32
2	B	195/207 (94%)	186 (95%)	9 (5%)	27 61
All	All	305/328 (93%)	285 (93%)	20 (7%)	16 49

5 of 20 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	170	GLU
2	B	216	ILE
2	B	290	LEU
2	B	265	GLU
1	A	418	ASN

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)

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

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

EDS was not executed - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates

EDS was not executed - this section is therefore empty.

6.4 Ligands

EDS was not executed - this section is therefore empty.

6.5 Other polymers

EDS was not executed - this section is therefore empty.