wwPDB X-ray Structure Validation Summary Report (i)

Dec 13, 2023 – 05:45 PM JST

PDB ID	:	8HR0
Title	:	The complex structure of COPII coat with HCoV-OC43 DD sorting motif
Authors	:	Ma, W.F.; Nan, Y.N.; Yang, M.R.; Li, Y.Q.
Deposited on	:	2022-12-14
Resolution	:	3.34 Å(reported)

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

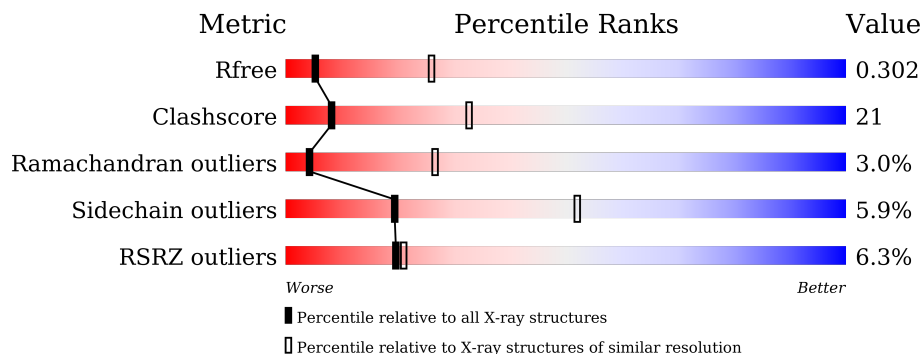|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 3.34 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		0.302
Clashscore		21
Ramachandran outliers		3.0%
Sidechain outliers		5.9%
RSRZ outliers		6.3%

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	1060 (3.38-3.30)
Clashscore	141614	1111 (3.38-3.30)
Ramachandran outliers	138981	1090 (3.38-3.30)
Sidechain outliers	138945	1089 (3.38-3.30)
RSRZ outliers	127900	1028 (3.38-3.30)

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	765	6% / 56% 33% · 7%
2	B	751	3% / 54% 39% · ·
3	C	199	15% / 36% 27% · · 32%
4	D	10	20% / 10% 10% 10% 70%

2 Entity composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 12401 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 Protein transport protein Sec23A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	708	Total	C	N	O	S	0	0	0
			5594	3569	964	1023	38			

- Molecule 2 is a protein called Protein transport protein Sec24A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	729	Total	C	N	O	S	0	0	0
			5717	3651	971	1062	33			

- Molecule 3 is a protein called Vesicle-trafficking protein SEC22b.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	135	Total	C	N	O	S	0	0	0
			1066	685	172	201	8			

- Molecule 4 is a protein called HCoV-OC43.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
4	D	3	Total	C	N	O	0	0	0
			22	11	3	8			

- Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	1	Total	Zn	0	0
			1	1		
5	B	1	Total	Zn	0	0
			1	1		

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: Protein transport protein Sec23A
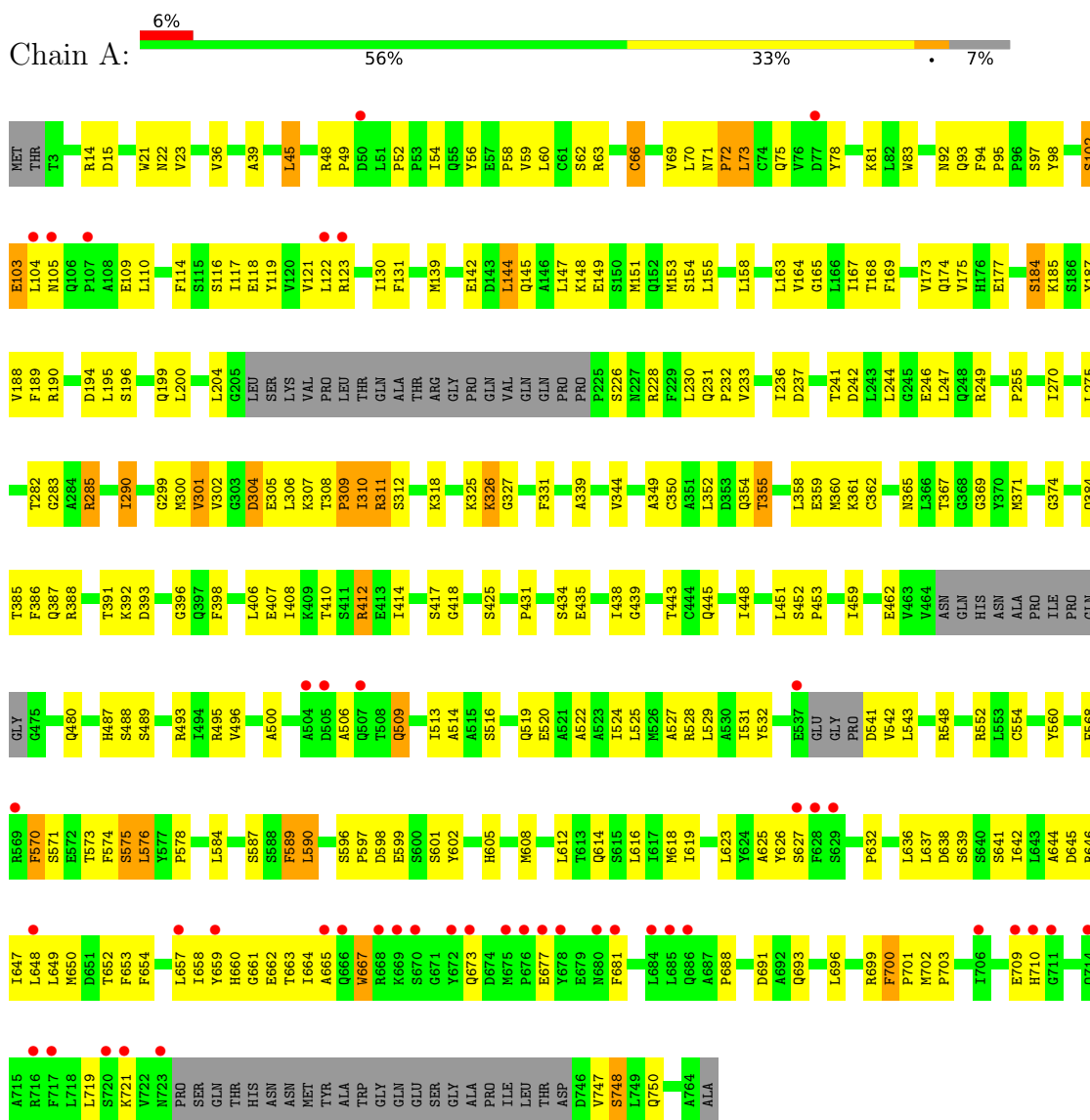
• Molecule 2: Protein transport protein Sec24A

● Molecule 3: Vesicle-trafficking protein SEC22b

Chain C:

● Molecule 4: HCoV-OC43

Chain D:

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	148.58Å 96.20Å 130.47Å 90.00° 90.01° 90.00°	Depositor
Resolution (Å)	49.02 – 3.34 49.02 – 3.34	Depositor EDS
% Data completeness (in resolution range)	98.4 (49.02-3.34) 98.4 (49.02-3.34)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.31 (at 3.33Å)	Xtriage
Refinement program	PHENIX 1.16.3549	Depositor
R, R_{free}	0.223 , 0.299 0.229 , 0.302	Depositor DCC
R_{free} test set	1994 reflections (7.53%)	wwPDB-VP
Wilson B-factor (Å^2)	73.2	Xtriage
Anisotropy	0.270	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 70.0	EDS
L-test for twinning[2]	$< |L| > = 0.46, < L^2 > = 0.29$	Xtriage
Estimated twinning fraction	0.029 for -h,-k,l	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	12401	wwPDB-VP
Average B, all atoms (Å^2)	84.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.96% 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: ZN

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	0.66	1/5725 (0.0%)	0.78	6/7753 (0.1%)
2	B	0.58	0/5839	0.78	3/7940 (0.0%)
3	C	0.45	0/1085	0.68	0/1465
4	D	0.76	0/21	2.22	1/26 (3.8%)
All	All	0.61	1/12670 (0.0%)	0.78	10/17184 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	435	GLU	CG-CD	5.58	1.60	1.51

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	285	ARG	CB-CA-C	-5.90	98.59	110.40
2	B	385	LEU	CB-CG-CD2	-5.74	101.25	111.00
1	A	45	LEU	CA-CB-CG	5.47	127.89	115.30
4	D	529	ASP	CB-CA-C	-5.36	99.69	110.40
1	A	285	ARG	CG-CD-NE	-5.27	100.74	111.80

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.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	5594	0	5528	221	0
2	B	5717	0	5737	253	0
3	C	1066	0	1045	59	0
4	D	22	0	9	6	0
5	A	1	0	0	0	0
5	B	1	0	0	0	0
All	All	12401	0	12319	525	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 21.

The worst 5 of 525 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:B:454:LEU:HD23	2:B:752:ARG:HG2	1.25	1.17
1:A:644:ALA:O	1:A:664:ILE:HD13	1.45	1.16
2:B:454:LEU:CD2	2:B:752:ARG:HG2	1.76	1.15
1:A:300:MET:N	1:A:326:LYS:HZ3	1.60	0.98
1:A:300:MET:H	1:A:326:LYS:HZ3	1.11	0.98

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	698/765 (91%)	621 (89%)	61 (9%)	16 (2%)	6	32
2	B	721/751 (96%)	619 (86%)	80 (11%)	22 (3%)	4	26
3	C	129/199 (65%)	100 (78%)	20 (16%)	9 (7%)	1	9
4	D	1/10 (10%)	1 (100%)	0	0	100	100
All	All	1549/1725 (90%)	1341 (87%)	161 (10%)	47 (3%)	4	27

5 of 47 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	575	SER
1	A	637	LEU
1	A	649	LEU
2	B	404	LYS
2	B	463	GLU

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	609/666 (91%)	577 (95%)	32 (5%)	22	56
2	B	647/682 (95%)	610 (94%)	37 (6%)	20	53
3	C	114/172 (66%)	104 (91%)	10 (9%)	10	36
4	D	2/10 (20%)	0	2 (100%)	0	0
All	All	1372/1530 (90%)	1291 (94%)	81 (6%)	19	52

5 of 81 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	777	PRO
3	C	35	SER
2	B	789	MET
2	B	994	LEU
3	C	80	PRO

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	532	ASN
2	B	736	GLN
2	B	932	ASN
2	B	765	ASN
2	B	447	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 ⓘ

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

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 ⓘ

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	708/765 (92%)	0.33	44 (6%)	20	21	42, 69, 171, 248	0
2	B	729/751 (97%)	0.12	24 (3%)	46	45	39, 65, 173, 252	0
3	C	135/199 (67%)	1.05	30 (22%)	0	0	69, 131, 198, 243	0
4	D	3/10 (30%)	2.66	2 (66%)	0	0	21, 21, 23, 30	0
All	All	1575/1725 (91%)	0.30	100 (6%)	20	21	21, 69, 177, 252	0

The worst 5 of 100 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	684	LEU	8.2
1	A	685	LEU	7.3
1	A	675	MET	6.7
1	A	680	ASN	5.9
2	B	1054	VAL	5.1

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
5	ZN	B	1101	1/1	0.93	0.18	76,76,76,76	0
5	ZN	A	801	1/1	0.99	0.07	81,81,81,81	0

6.5 Other polymers ⓘ

There are no such residues in this entry.