wwPDB X-ray Structure Validation Summary Report (i)

Oct 9, 2023 – 04:20 PM EDT

PDB ID	:	8DRY
Title	:	Product structure of SARS-CoV-2 Mpro C145A mutant in complex with nsp12-nsp13 (C12) cut site sequence
Authors	:	Lee, J.; Kenward, C.; Worrall, L.J.; Vuckovic, M.; Paetzel, M.; Strynadka, N.C.J.
Deposited on	:	2022-07-21
Resolution	:	2.49 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

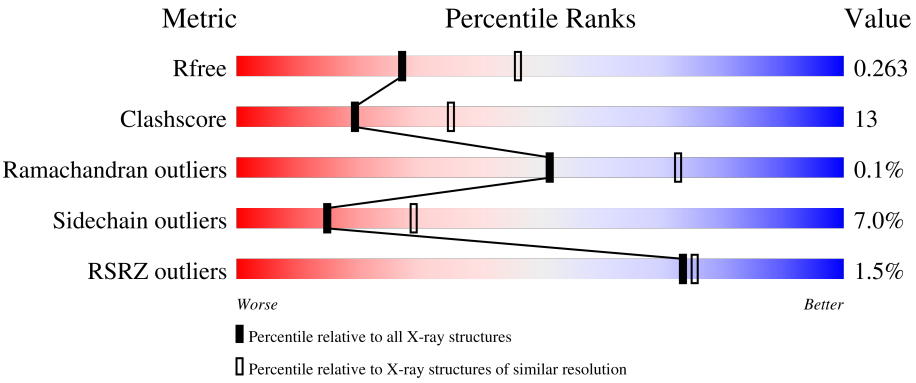
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.49 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	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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 <=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	306	75% 23%
1	B	306	72% 26%
1	C	306	71% 27%
1	D	306	63% 34%
1	E	306	72% 26%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	306	2% 71% 27%
1	G	306	% 66% 32%
1	H	306	60% 36%
1	I	306	4% 64% 32%
1	J	306	8% 58% 38% 5%
1	K	306	75% 22%
1	L	306	% 76% 22%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 28308 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 Fusion protein of 3C-like proteinase nsp5 and nsp12-nsp13 (C12) cut site.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	303	Total	C	N	O	S	0	0	0
			2342	1484	400	437	21			
1	B	303	Total	C	N	O	S	0	1	0
			2354	1491	403	439	21			
1	C	306	Total	C	N	O	S	0	0	0
			2366	1497	404	444	21			
1	D	306	Total	C	N	O	S	0	0	0
			2367	1500	404	442	21			
1	E	303	Total	C	N	O	S	0	1	0
			2350	1489	403	437	21			
1	F	302	Total	C	N	O	S	0	0	0
			2339	1482	399	437	21			
1	G	302	Total	C	N	O	S	0	0	0
			2328	1474	399	434	21			
1	H	302	Total	C	N	O	S	0	1	0
			2340	1482	399	438	21			
1	I	306	Total	C	N	O	S	0	0	0
			2352	1490	399	442	21			
1	J	306	Total	C	N	O	S	0	0	0
			2329	1474	398	437	20			
1	K	303	Total	C	N	O	S	0	0	0
			2329	1474	400	434	21			
1	L	302	Total	C	N	O	S	0	0	0
			2339	1482	399	437	21			

There are 12 discrepancies between the modelled and reference sequences:

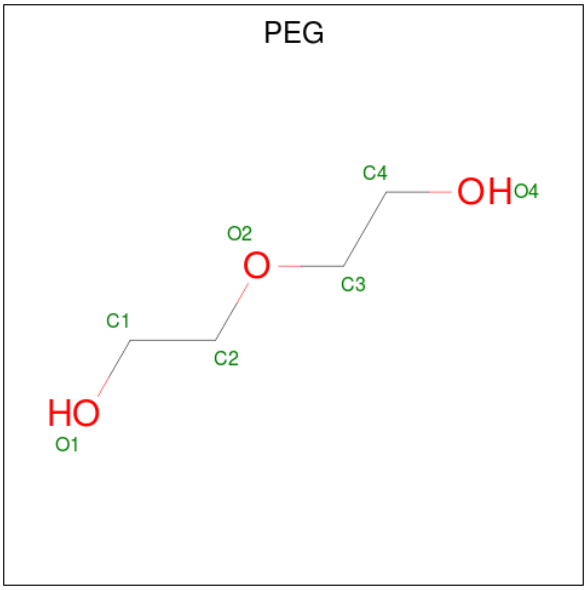
Chain	Residue	Modelled	Actual	Comment	Reference
A	145	ALA	CYS	engineered mutation	UNP P0DTD1
B	145	ALA	CYS	engineered mutation	UNP P0DTD1
C	145	ALA	CYS	engineered mutation	UNP P0DTD1
D	145	ALA	CYS	engineered mutation	UNP P0DTD1

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	145	ALA	CYS	engineered mutation	UNP P0DTD1
F	145	ALA	CYS	engineered mutation	UNP P0DTD1
G	145	ALA	CYS	engineered mutation	UNP P0DTD1
H	145	ALA	CYS	engineered mutation	UNP P0DTD1
I	145	ALA	CYS	engineered mutation	UNP P0DTD1
J	145	ALA	CYS	engineered mutation	UNP P0DTD1
K	145	ALA	CYS	engineered mutation	UNP P0DTD1
L	145	ALA	CYS	engineered mutation	UNP P0DTD1

- Molecule 2 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	1	Total C O 7 4 3	0	0
2	E	1	Total C O 7 4 3	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	23	Total O 23 23	0	0
3	B	28	Total O 28 28	0	0
3	C	13	Total O 13 13	0	0

Continued on next page...

Continued from previous page...

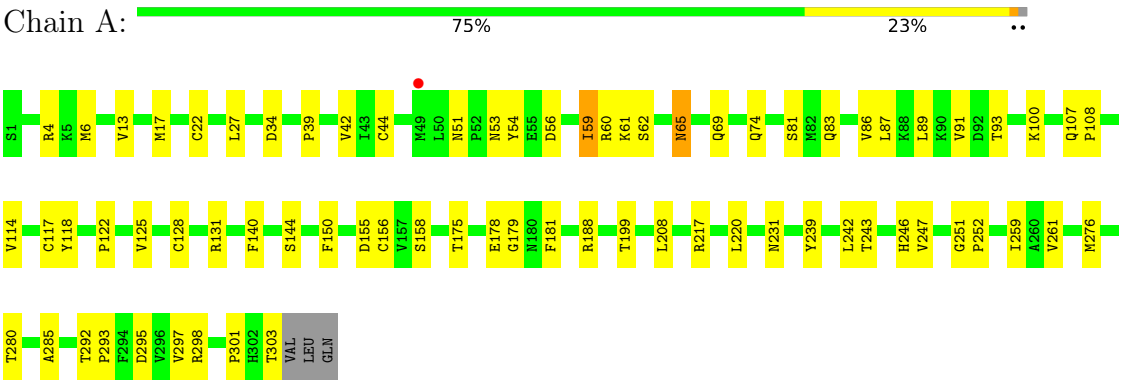
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	D	6	Total O 6 6	0	0
3	E	20	Total O 20 20	0	0
3	F	17	Total O 17 17	0	0
3	G	12	Total O 12 12	0	0
3	H	5	Total O 5 5	0	0
3	I	1	Total O 1 1	0	0
3	J	2	Total O 2 2	0	0
3	K	15	Total O 15 15	0	0
3	L	17	Total O 17 17	0	0

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.

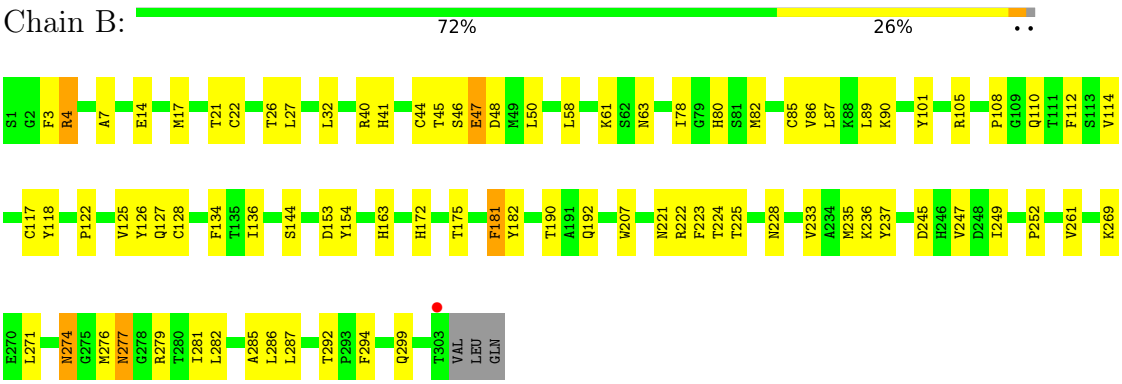
• Molecule 1: Fusion protein of 3C-like proteinase nsp5 and nsp12-nsp13 (C12) cut site

Chain A:



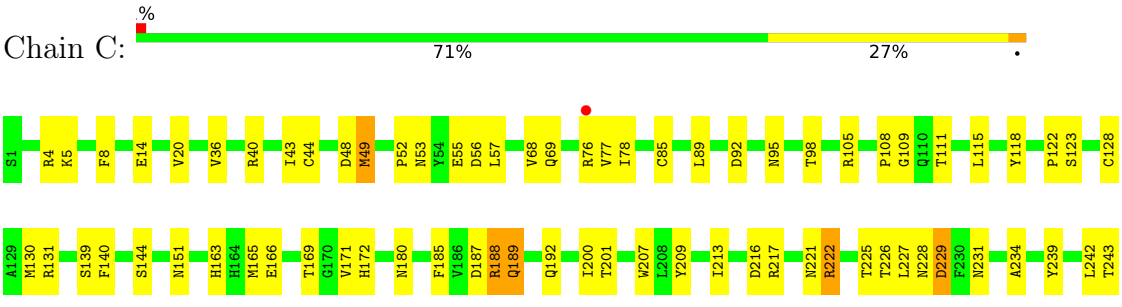
• Molecule 1: Fusion protein of 3C-like proteinase nsp5 and nsp12-nsp13 (C12) cut site

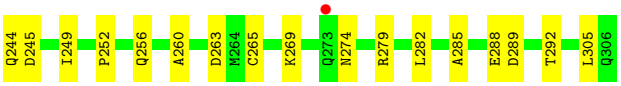
Chain B:



• Molecule 1: Fusion protein of 3C-like proteinase nsp5 and nsp12-nsp13 (C12) cut site

Chain C:



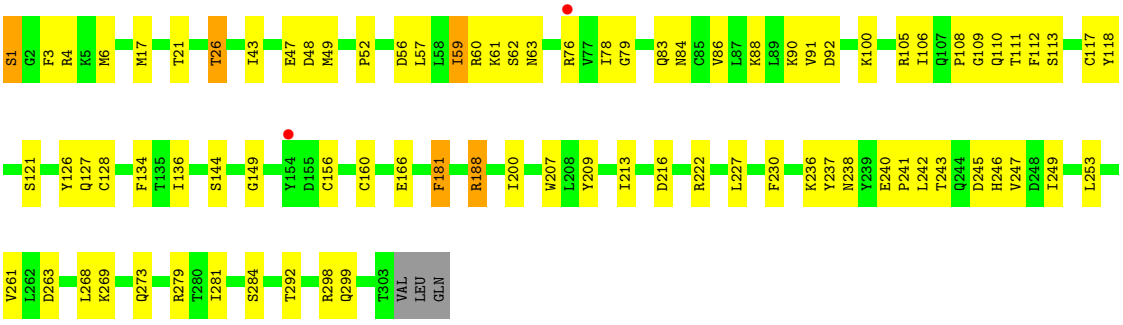


• Molecule 1: Fusion protein of 3C-like proteinase nsp5 and nsp12-nsp13 (C12) cut site

Chain D:

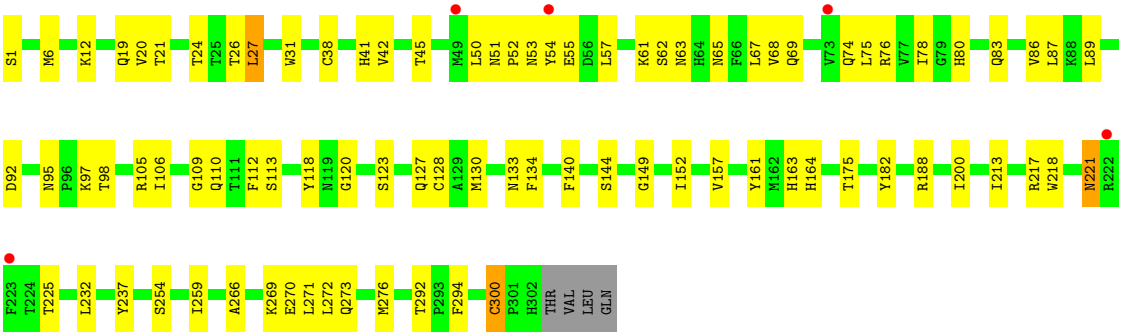
• Molecule 1: Fusion protein of 3C-like proteinase nsp5 and nsp12-nsp13 (C12) cut site

Chain E:



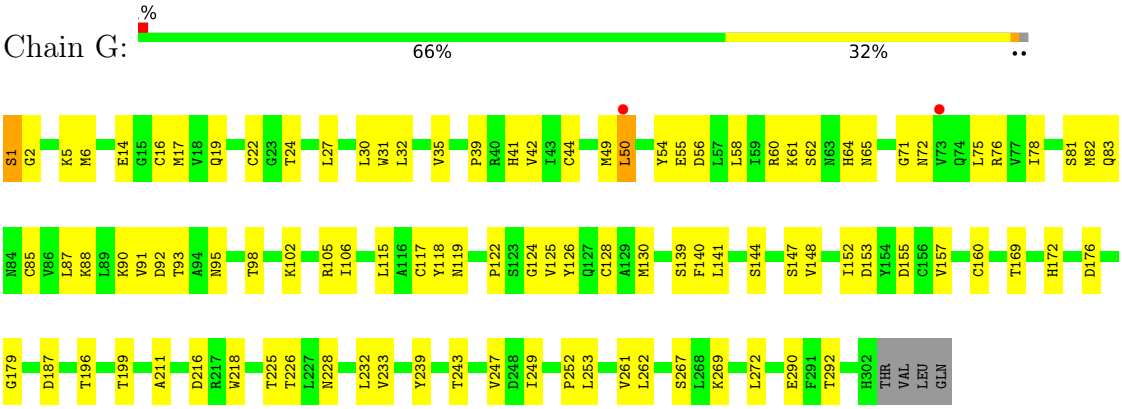
• Molecule 1: Fusion protein of 3C-like proteinase nsp5 and nsp12-nsp13 (C12) cut site

Chain F:



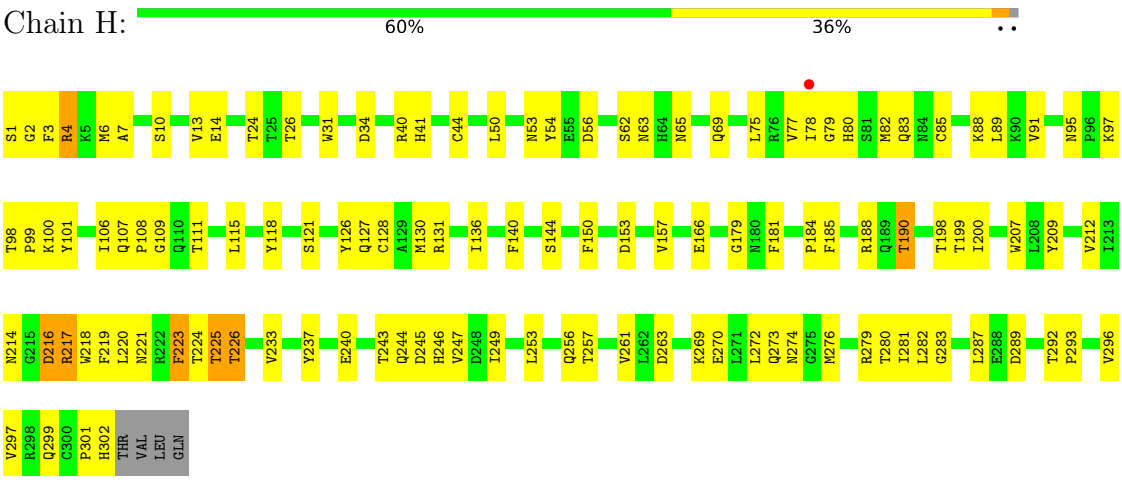
• Molecule 1: Fusion protein of 3C-like proteinase nsp5 and nsp12-nsp13 (C12) cut site

Chain G:



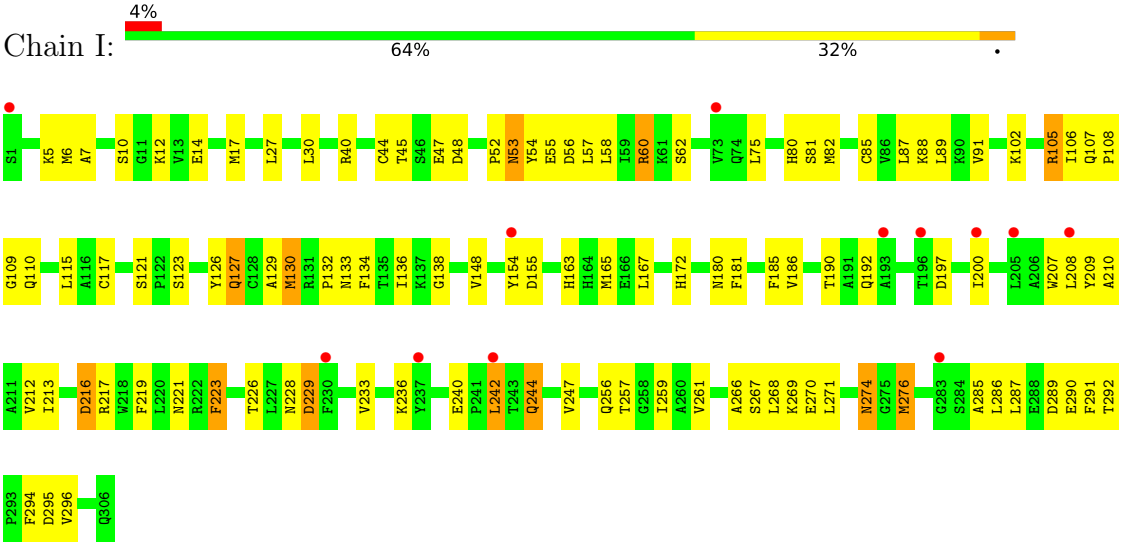
• Molecule 1: Fusion protein of 3C-like proteinase nsp5 and nsp12-nsp13 (C12) cut site

Chain H:



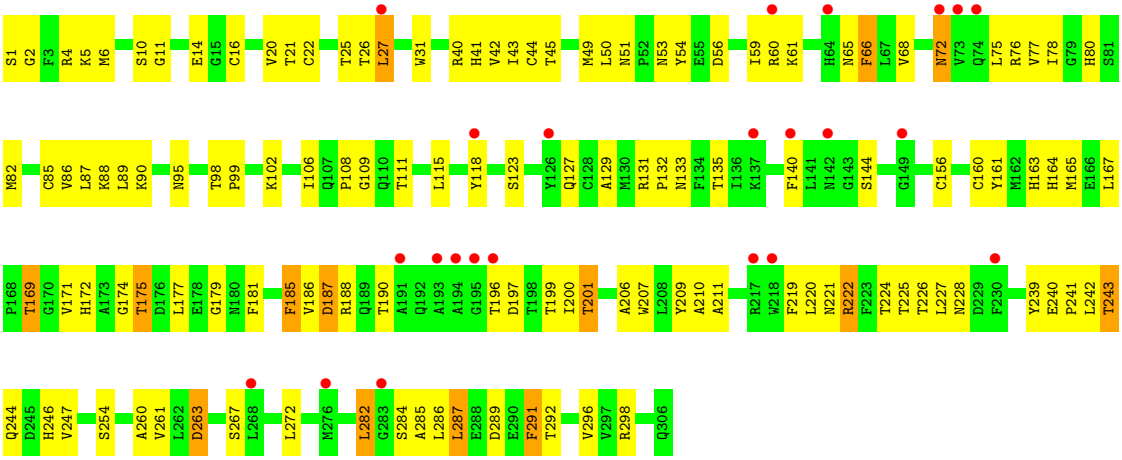
• Molecule 1: Fusion protein of 3C-like proteinase nsp5 and nsp12-nsp13 (C12) cut site

Chain I:



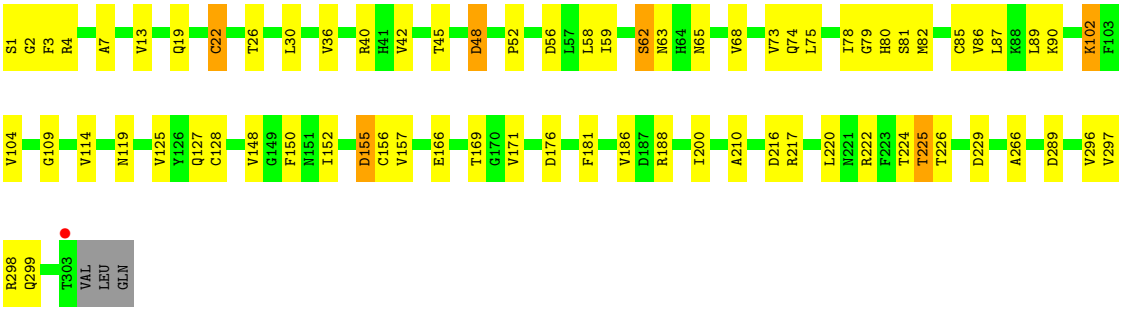
• Molecule 1: Fusion protein of 3C-like proteinase nsp5 and nsp12-nsp13 (C12) cut site

Chain J:



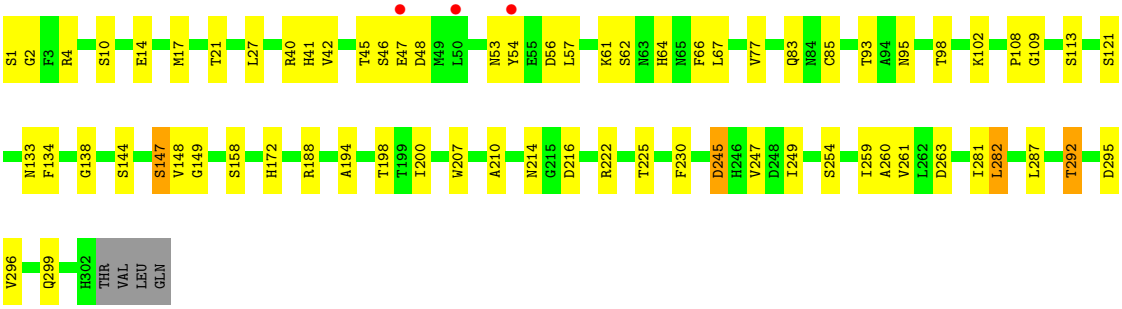
• Molecule 1: Fusion protein of 3C-like proteinase nsp5 and nsp12-nsp13 (C12) cut site

Chain K:



• Molecule 1: Fusion protein of 3C-like proteinase nsp5 and nsp12-nsp13 (C12) cut site

Chain L:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	67.33Å 107.92Å 277.25Å 90.00° 90.73° 90.00°	Depositor
Resolution (Å)	65.23 – 2.49 65.23 – 2.49	Depositor EDS
% Data completeness (in resolution range)	72.5 (65.23-2.49) 72.6 (65.23-2.49)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.02 (at 2.48Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.194 , 0.264 0.195 , 0.263	Depositor DCC
R_{free} test set	2006 reflections (1.99%)	wwPDB-VP
Wilson B-factor (Å^2)	51.6	Xtriage
Anisotropy	0.105	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.24 , 27.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.053 for h,-k,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	28308	wwPDB-VP
Average B, all atoms (Å^2)	59.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 23.60 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 4.5167e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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: PEG

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	0.53	0/2396	0.68	0/3259
1	B	0.56	0/2411	0.71	0/3278
1	C	0.48	0/2420	0.64	0/3290
1	D	0.49	0/2421	0.66	0/3292
1	E	0.52	0/2407	0.68	0/3273
1	F	0.48	0/2393	0.66	0/3254
1	G	0.46	0/2381	0.63	0/3238
1	H	0.46	0/2397	0.66	0/3261
1	I	0.41	0/2406	0.63	0/3273
1	J	0.42	0/2382	0.64	0/3244
1	K	0.49	0/2382	0.69	0/3240
1	L	0.51	0/2393	0.68	0/3254
All	All	0.48	0/28789	0.66	0/39156

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	2342	0	2290	49	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	2354	0	2307	54	0
1	C	2366	0	2306	56	0
1	D	2367	0	2318	76	0
1	E	2350	0	2303	53	0
1	F	2339	0	2287	58	0
1	G	2328	0	2276	67	0
1	H	2340	0	2282	91	0
1	I	2352	0	2283	74	0
1	J	2329	0	2242	96	0
1	K	2329	0	2272	46	0
1	L	2339	0	2287	40	0
2	B	7	0	10	0	0
2	E	7	0	10	0	0
3	A	23	0	0	1	0
3	B	28	0	0	2	0
3	C	13	0	0	0	0
3	D	6	0	0	0	0
3	E	20	0	0	3	0
3	F	17	0	0	2	0
3	G	12	0	0	1	0
3	H	5	0	0	1	0
3	I	1	0	0	0	0
3	J	2	0	0	0	0
3	K	15	0	0	1	0
3	L	17	0	0	1	0
All	All	28308	0	27473	712	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 13.

The worst 5 of 712 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:J:5:LYS:HE3	1:J:291:PHE:HE1	1.34	0.90
1:J:161:TYR:HE1	1:J:174:GLY:HA3	1.38	0.87
1:H:109:GLY:HA2	1:H:200:ILE:HD13	1.57	0.86
1:B:4:ARG:H	1:B:299:GLN:HE22	1.20	0.86
1:H:106:ILE:HD11	1:H:130:MET:HE3	1.59	0.85

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	301/306 (98%)	289 (96%)	12 (4%)	0	100	100
1	B	302/306 (99%)	289 (96%)	13 (4%)	0	100	100
1	C	304/306 (99%)	285 (94%)	19 (6%)	0	100	100
1	D	304/306 (99%)	291 (96%)	12 (4%)	1 (0%)	41	61
1	E	302/306 (99%)	283 (94%)	19 (6%)	0	100	100
1	F	300/306 (98%)	277 (92%)	23 (8%)	0	100	100
1	G	300/306 (98%)	288 (96%)	12 (4%)	0	100	100
1	H	301/306 (98%)	282 (94%)	18 (6%)	1 (0%)	41	61
1	I	304/306 (99%)	286 (94%)	18 (6%)	0	100	100
1	J	304/306 (99%)	283 (93%)	21 (7%)	0	100	100
1	K	301/306 (98%)	290 (96%)	11 (4%)	0	100	100
1	L	300/306 (98%)	283 (94%)	17 (6%)	0	100	100
All	All	3623/3672 (99%)	3426 (95%)	195 (5%)	2 (0%)	51	73

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	H	282	LEU
1	D	154	TYR

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	A	259/263 (98%)	251 (97%)	8 (3%)	40 67
1	B	261/263 (99%)	243 (93%)	18 (7%)	15 30
1	C	261/263 (99%)	245 (94%)	16 (6%)	18 36
1	D	262/263 (100%)	242 (92%)	20 (8%)	13 25
1	E	260/263 (99%)	242 (93%)	18 (7%)	15 30
1	F	259/263 (98%)	249 (96%)	10 (4%)	32 57
1	G	257/263 (98%)	241 (94%)	16 (6%)	18 35
1	H	259/263 (98%)	242 (93%)	17 (7%)	16 32
1	I	258/263 (98%)	231 (90%)	27 (10%)	7 13
1	J	252/263 (96%)	218 (86%)	34 (14%)	4 7
1	K	256/263 (97%)	237 (93%)	19 (7%)	13 27
1	L	259/263 (98%)	246 (95%)	13 (5%)	24 46
All	All	3103/3156 (98%)	2887 (93%)	216 (7%)	15 29

5 of 216 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	H	190	THR
1	I	233	VAL
1	K	225	THR
1	H	223	PHE
1	I	105	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 79 such sidechains are listed below:

Mol	Chain	Res	Type
1	I	172	HIS
1	L	41	HIS
1	I	246	HIS
1	J	110	GLN
1	L	119	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)

2 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	PEG	B	401	-	6,6,6	0.24	0	5,5,5	0.23	0
2	PEG	E	401	-	6,6,6	0.35	0	5,5,5	0.20	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	PEG	B	401	-	-	2/4/4/4	-
2	PEG	E	401	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	E	401	PEG	O2-C3-C4-O4
2	B	401	PEG	O1-C1-C2-O2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
2	E	401	PEG	C1-C2-O2-C3
2	B	401	PEG	C1-C2-O2-C3

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, 95^{th} 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	303/306 (99%)	-0.49	1 (0%)	94	94	23, 38, 77, 131	0
1	B	303/306 (99%)	-0.52	1 (0%)	94	94	21, 37, 64, 114	0
1	C	306/306 (100%)	-0.30	2 (0%)	87	89	28, 54, 98, 144	0
1	D	306/306 (100%)	-0.39	0	100	100	30, 52, 102, 129	0
1	E	303/306 (99%)	-0.50	2 (0%)	87	89	24, 42, 76, 145	0
1	F	302/306 (98%)	-0.42	5 (1%)	70	72	28, 48, 91, 149	0
1	G	302/306 (98%)	-0.43	2 (0%)	87	89	38, 61, 94, 123	0
1	H	302/306 (98%)	-0.25	1 (0%)	94	94	42, 65, 102, 122	0
1	I	306/306 (100%)	0.11	12 (3%)	39	42	44, 84, 127, 156	0
1	J	306/306 (100%)	0.30	23 (7%)	14	14	35, 86, 130, 181	0
1	K	303/306 (99%)	-0.51	1 (0%)	94	94	28, 44, 76, 110	0
1	L	302/306 (98%)	-0.40	3 (0%)	82	84	31, 50, 88, 155	0
All	All	3644/3672 (99%)	-0.32	53 (1%)	73	75	21, 53, 109, 181	0

The worst 5 of 53 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	I	154	TYR	5.3
1	J	194	ALA	4.7
1	J	126	TYR	4.4
1	J	196	THR	4.4
1	F	49	MET	3.9

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 (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

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, 95^{th} 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	PEG	E	401	7/7	0.79	0.18	46,50,54,57	0
2	PEG	B	401	7/7	0.90	0.14	44,47,55,55	0

6.5 Other polymers (i)

There are no such residues in this entry.