wwPDB X-ray Structure Validation Summary Report (i)

Jan 15, 2024 – 03:22 pm GMT

PDB ID	:	6QNW
Title	:	Influenza A Polymerase Heterotrimer Human H3N2 Northern Territory 1968
Authors	:	Keown, J.R.; Fan, H.; Grimes, J.M.
Deposited on	:	2019-02-12
Resolution	:	3.31 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 (i)) were used in the production of this report:

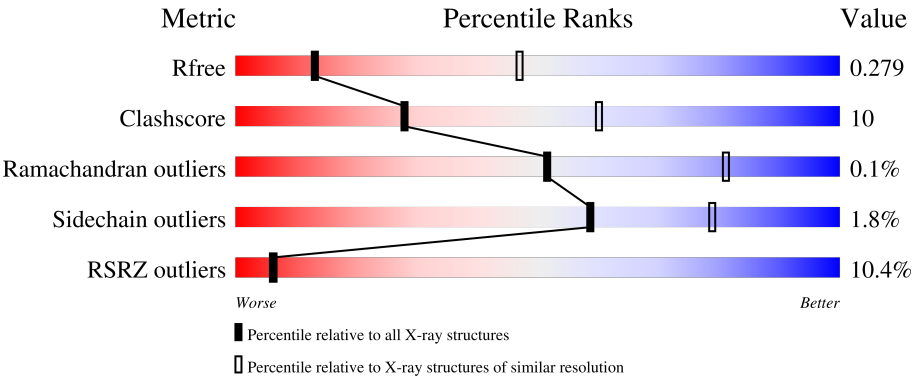
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation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

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 3.31 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.



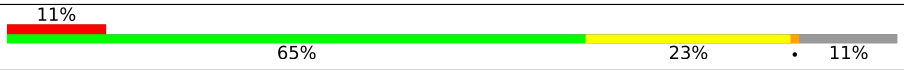
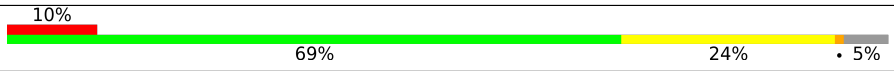
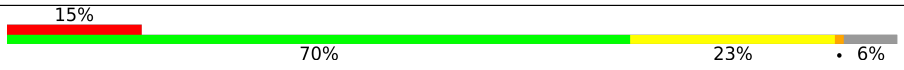
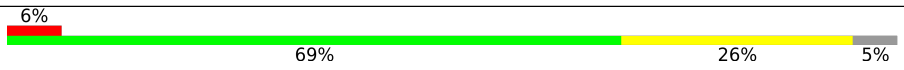
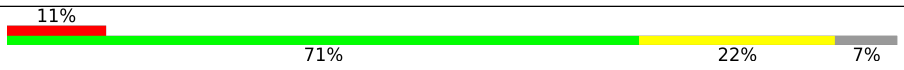
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1089 (3.36-3.28)
Clashscore	141614	1137 (3.36-3.28)
Ramachandran outliers	138981	1115 (3.36-3.28)
Sidechain outliers	138945	1114 (3.36-3.28)
RSRZ outliers	127900	1059 (3.36-3.28)

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	716	3% 72% 25%
1	D	716	17% 73% 21%
1	G	716	2% 72% 25%
1	J	716	10% 75% 22%
2	B	757	4% 70% 20% 10%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	E	757	
2	H	757	
2	K	757	
3	C	765	
3	F	765	
3	I	765	
3	L	765	

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 67232 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 Polymerase acidic protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	704	Total	C	N	O	S	0	0	0
			5733	3638	969	1084	42			
1	D	686	Total	C	N	O	S	0	0	0
			5604	3556	949	1058	41			
1	G	701	Total	C	N	O	S	0	0	0
			5707	3617	967	1081	42			
1	J	702	Total	C	N	O	S	0	0	0
			5713	3622	966	1083	42			

- Molecule 2 is a protein called RNA-directed RNA polymerase catalytic subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	679	Total	C	N	O	S	0	0	0
			5426	3417	932	1034	43			
2	E	665	Total	C	N	O	S	0	0	0
			5315	3351	910	1011	43			
2	H	676	Total	C	N	O	S	0	0	0
			5410	3410	929	1028	43			
2	K	673	Total	C	N	O	S	0	0	0
			5382	3390	925	1024	43			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	577	LYS	GLU	conflict	UNP P03432
E	577	LYS	GLU	conflict	UNP P03432
H	577	LYS	GLU	conflict	UNP P03432
K	577	LYS	GLU	conflict	UNP P03432

- Molecule 3 is a protein called Polymerase basic protein 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	724	Total	C	N	O	S	0	0	0
			5772	3627	1044	1060	41			
3	F	720	Total	C	N	O	S	0	0	0
			5734	3604	1034	1056	40			
3	I	726	Total	C	N	O	S	0	0	0
			5789	3638	1047	1063	41			
3	L	709	Total	C	N	O	S	0	0	0
			5647	3552	1015	1039	41			

There are 24 discrepancies between the modelled and reference sequences:

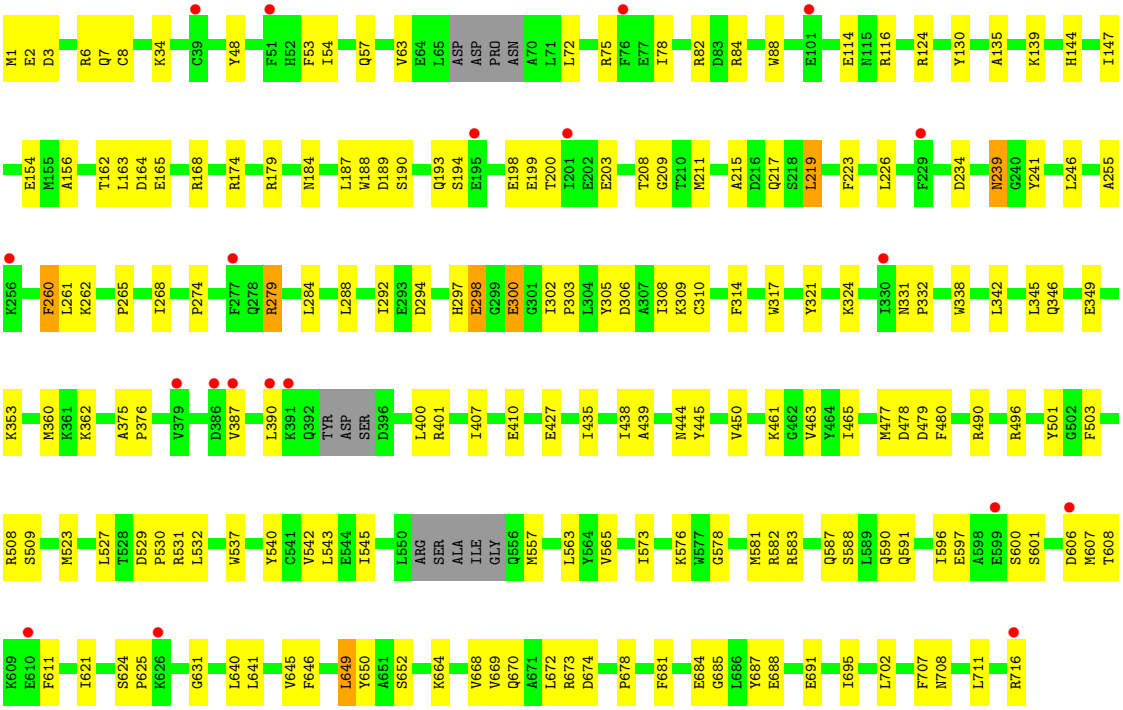
Chain	Residue	Modelled	Actual	Comment	Reference
C	760	GLU	-	expression tag	UNP P03429
C	761	ASN	-	expression tag	UNP P03429
C	762	LEU	-	expression tag	UNP P03429
C	763	TYR	-	expression tag	UNP P03429
C	764	PHE	-	expression tag	UNP P03429
C	765	GLN	-	expression tag	UNP P03429
F	760	GLU	-	expression tag	UNP P03429
F	761	ASN	-	expression tag	UNP P03429
F	762	LEU	-	expression tag	UNP P03429
F	763	TYR	-	expression tag	UNP P03429
F	764	PHE	-	expression tag	UNP P03429
F	765	GLN	-	expression tag	UNP P03429
I	760	GLU	-	expression tag	UNP P03429
I	761	ASN	-	expression tag	UNP P03429
I	762	LEU	-	expression tag	UNP P03429
I	763	TYR	-	expression tag	UNP P03429
I	764	PHE	-	expression tag	UNP P03429
I	765	GLN	-	expression tag	UNP P03429
L	760	GLU	-	expression tag	UNP P03429
L	761	ASN	-	expression tag	UNP P03429
L	762	LEU	-	expression tag	UNP P03429
L	763	TYR	-	expression tag	UNP P03429
L	764	PHE	-	expression tag	UNP P03429
L	765	GLN	-	expression tag	UNP P03429

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.

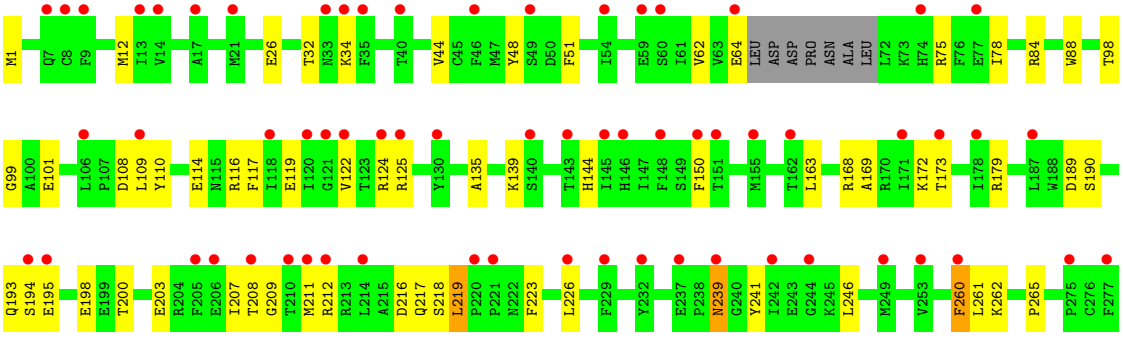
• Molecule 1: Polymerase acidic protein

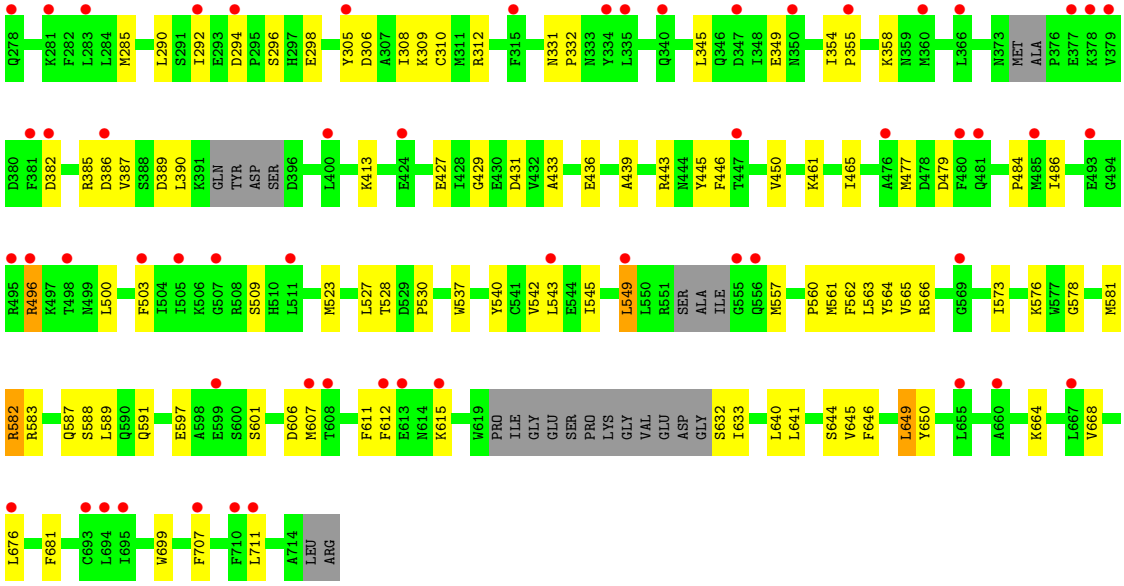
Chain A:



• Molecule 1: Polymerase acidic protein

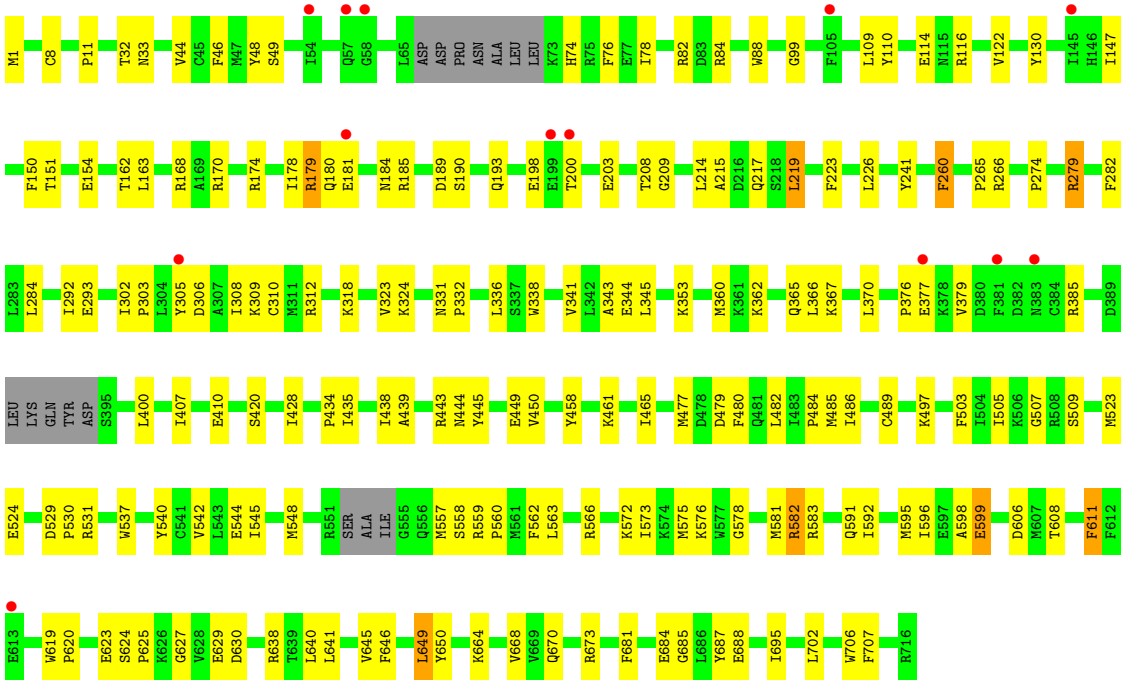
Chain D:





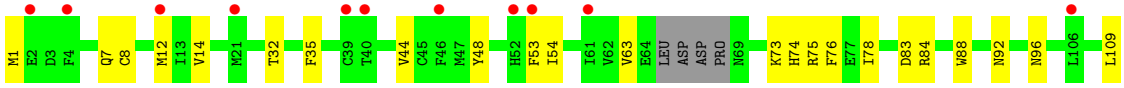
• Molecule 1: Polymerase acidic protein

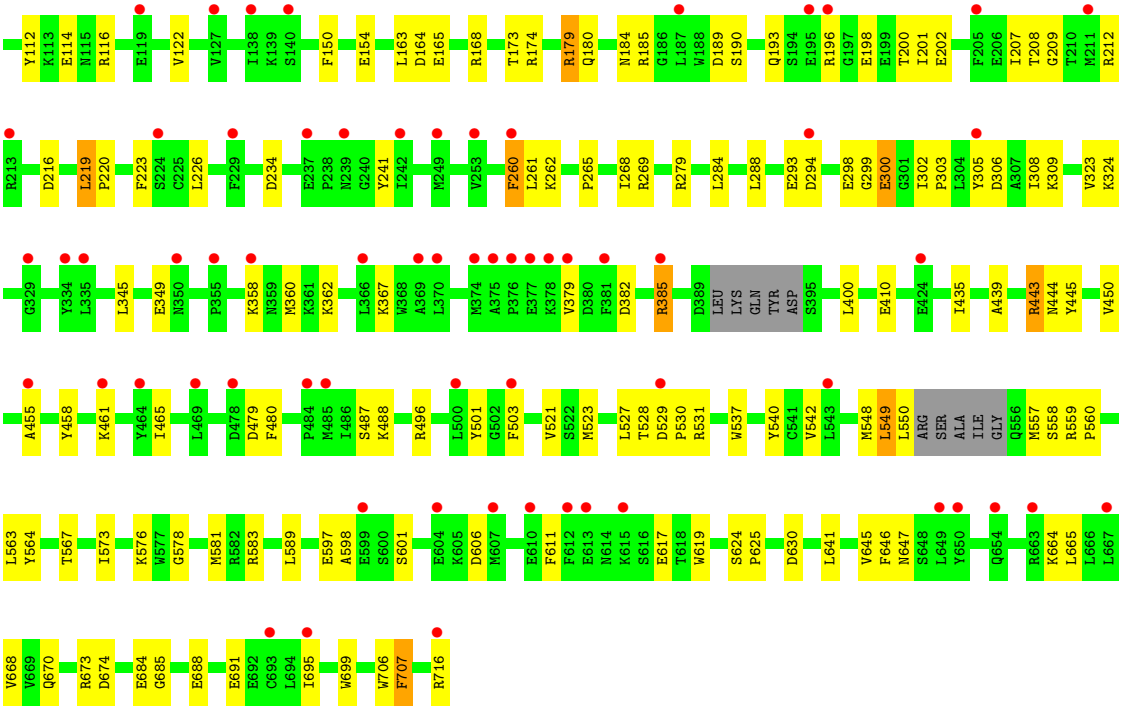
Chain G:



• Molecule 1: Polymerase acidic protein

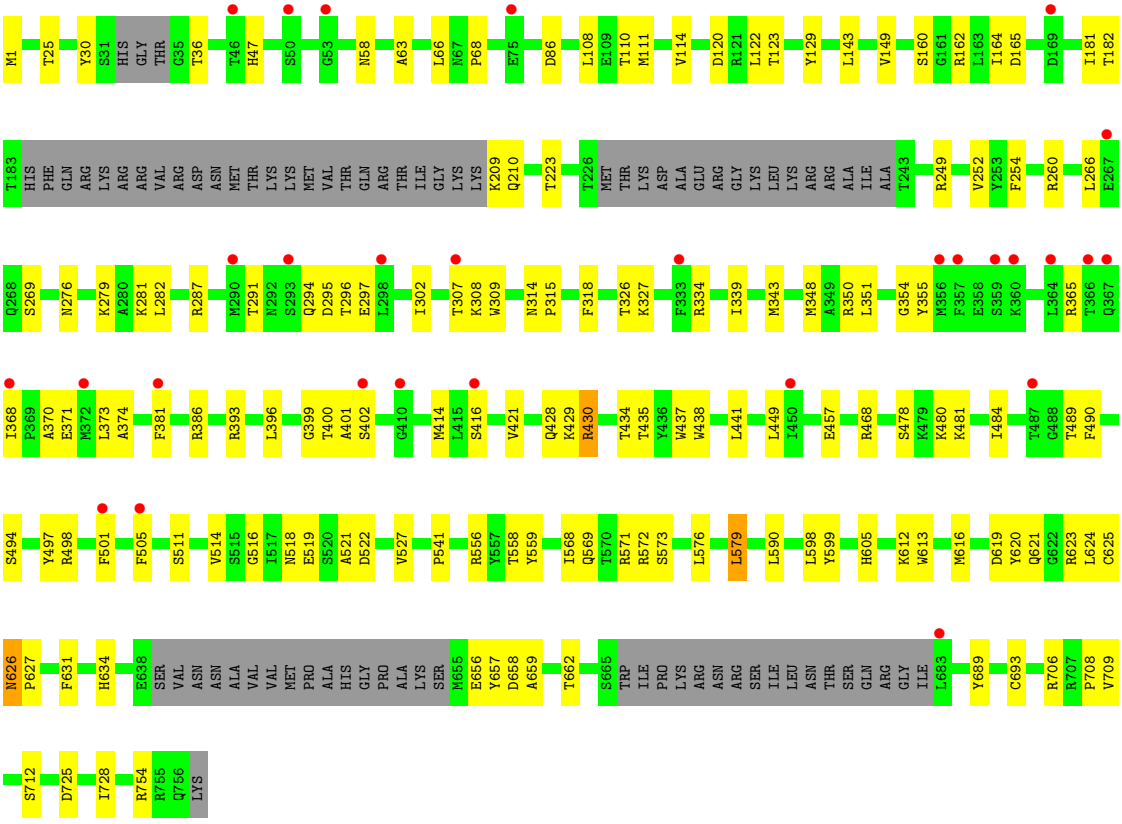
Chain J:





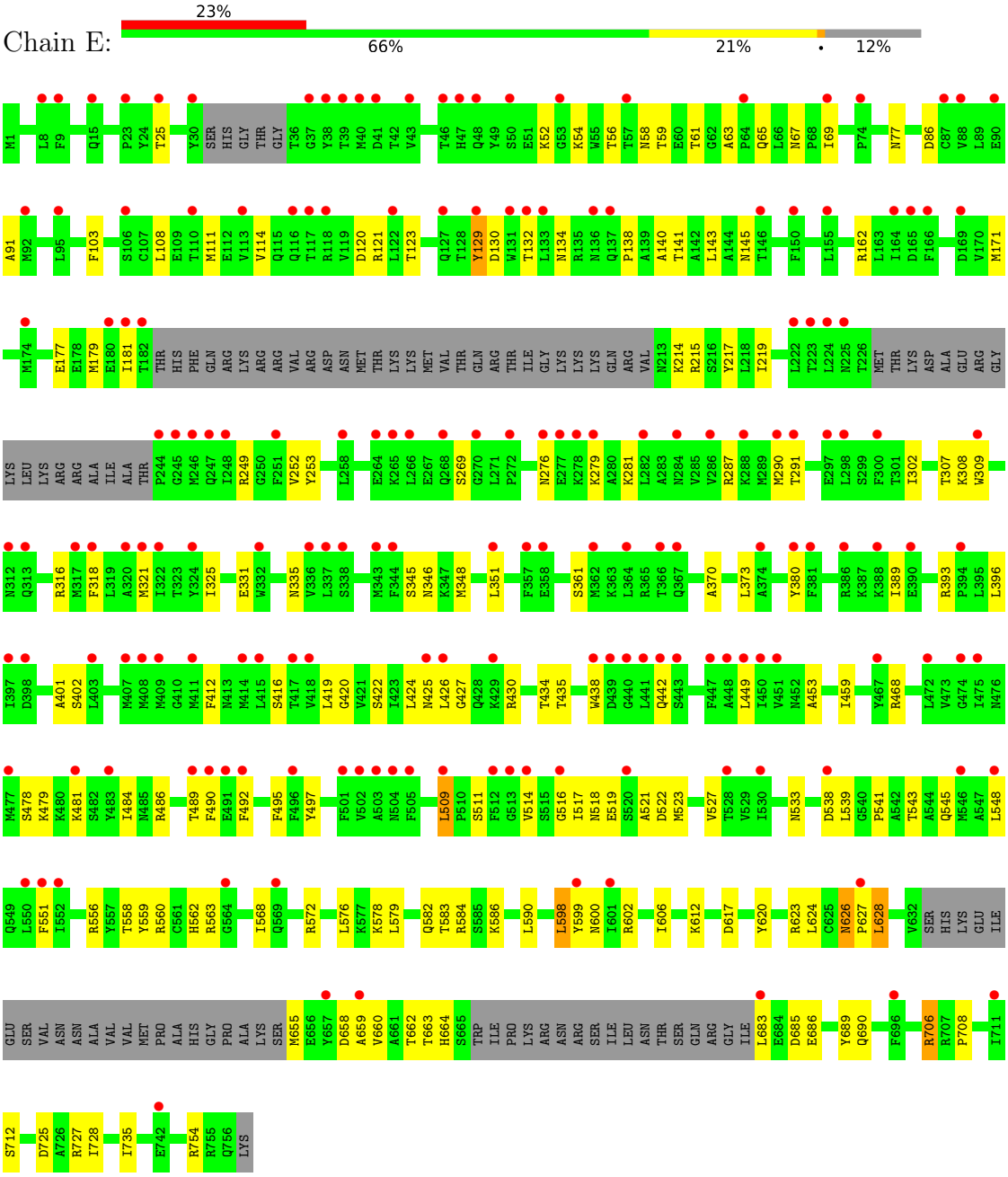
• Molecule 2: RNA-directed RNA polymerase catalytic subunit

Chain B:



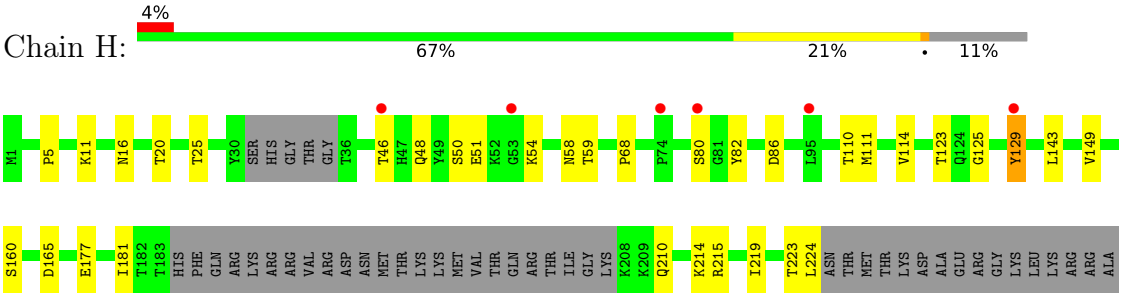
• Molecule 2: RNA-directed RNA polymerase catalytic subunit

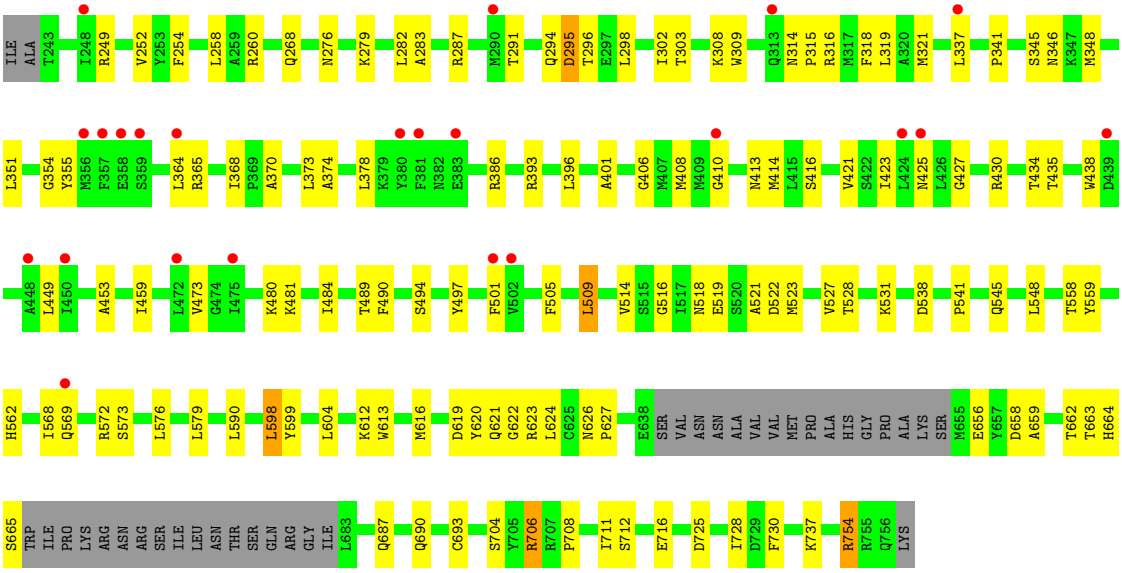
Chain E:



• Molecule 2: RNA-directed RNA polymerase catalytic subunit

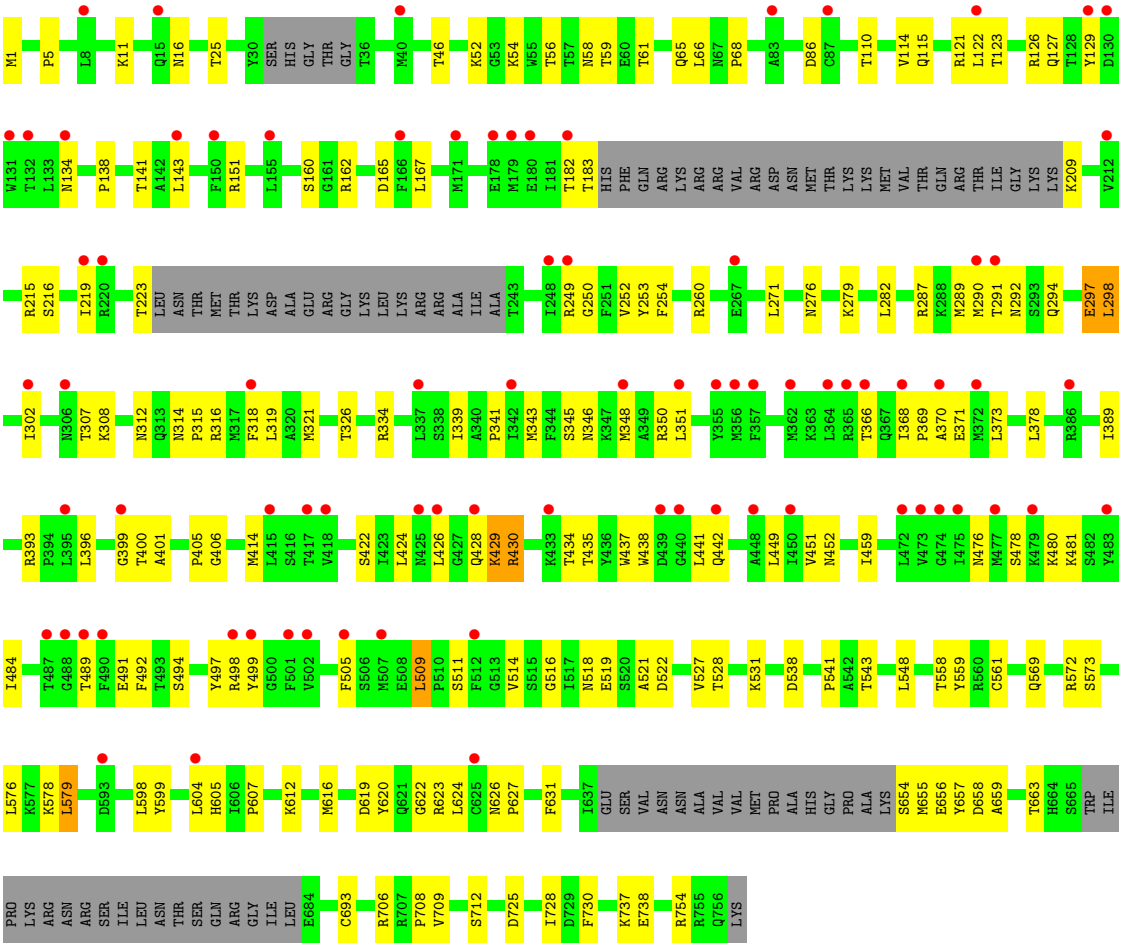
Chain H:





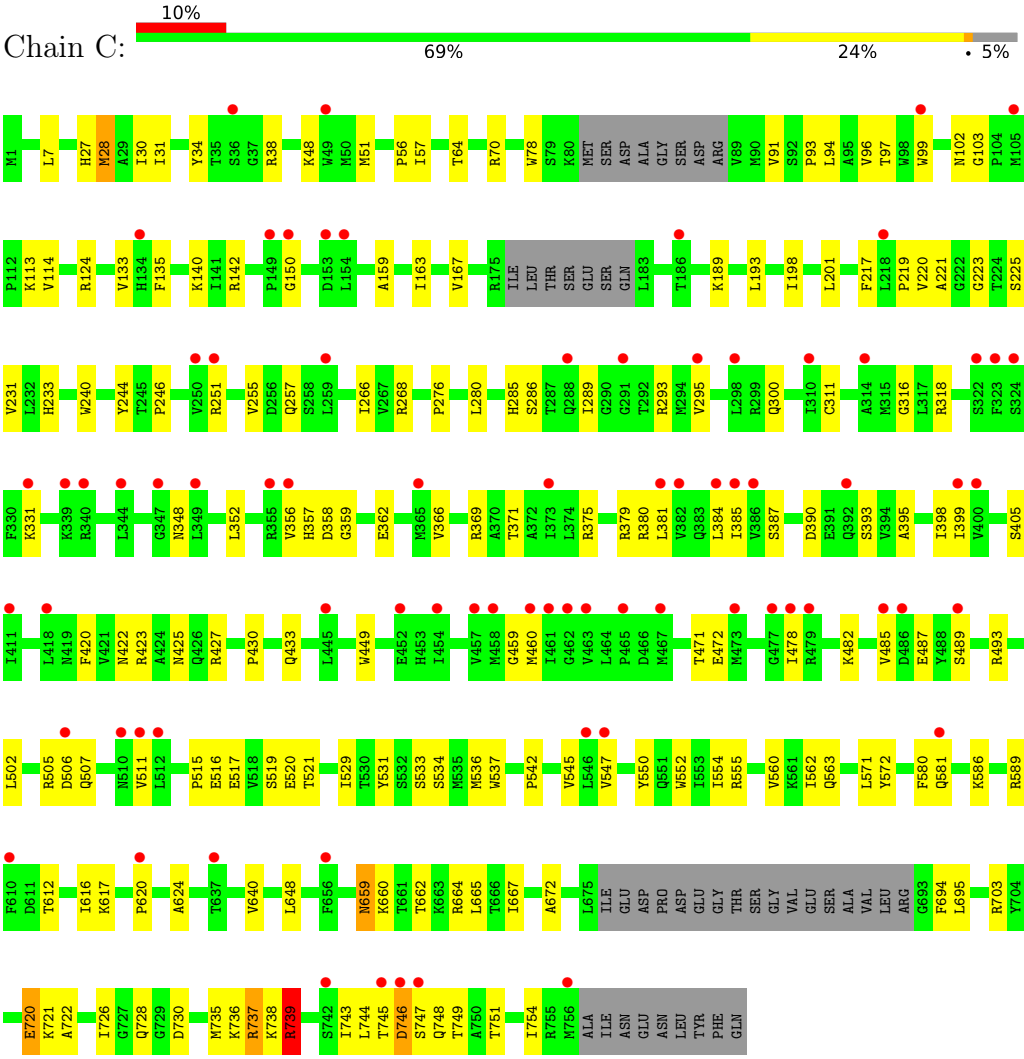
• Molecule 2: RNA-directed RNA polymerase catalytic subunit

Chain K:



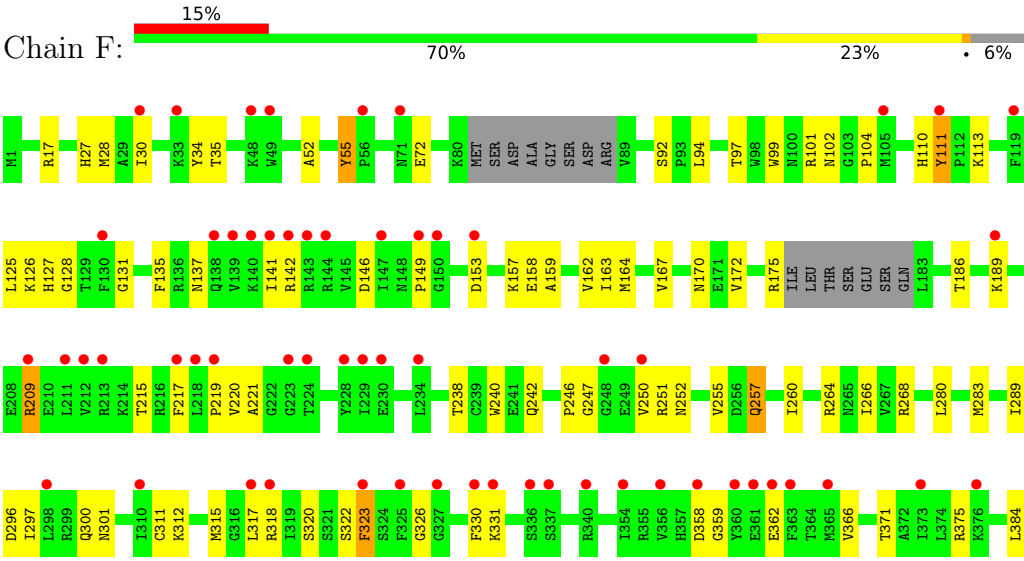
• Molecule 3: Polymerase basic protein 2

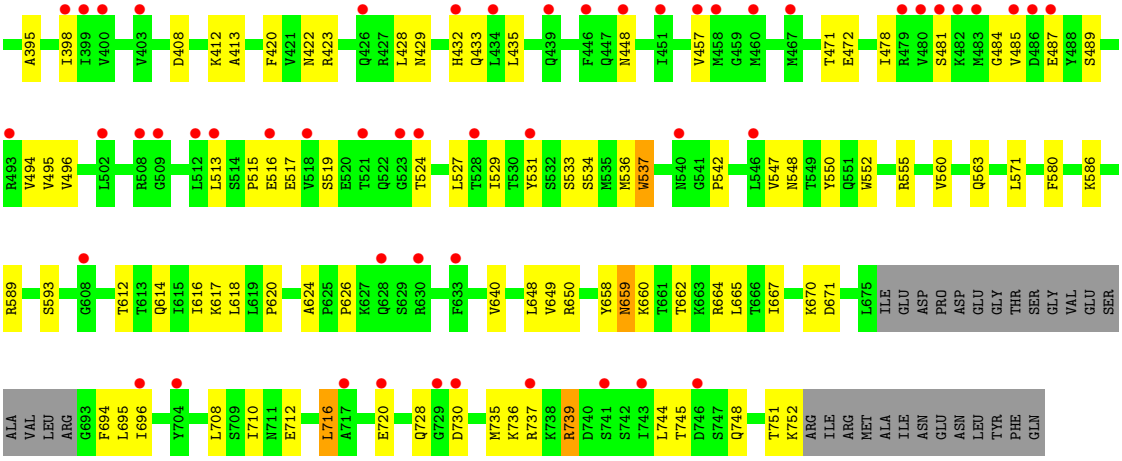
Chain C:



• Molecule 3: Polymerase basic protein 2

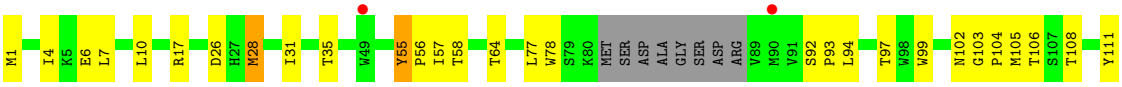
Chain F:

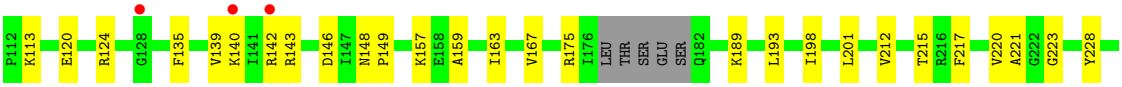


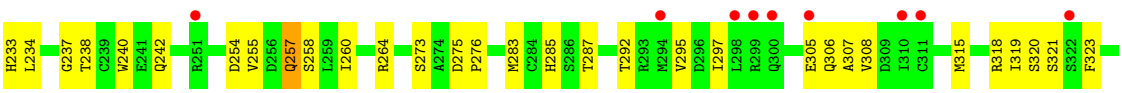


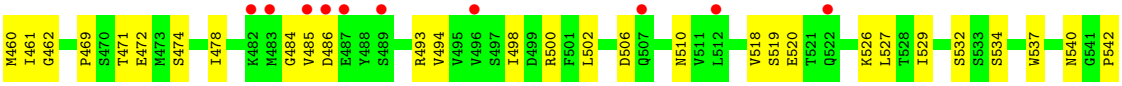
• Molecule 3: Polymerase basic protein 2

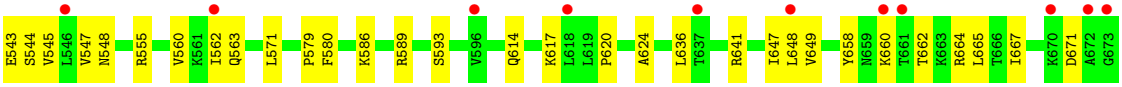
Chain I:

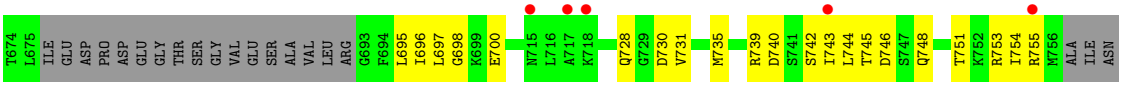






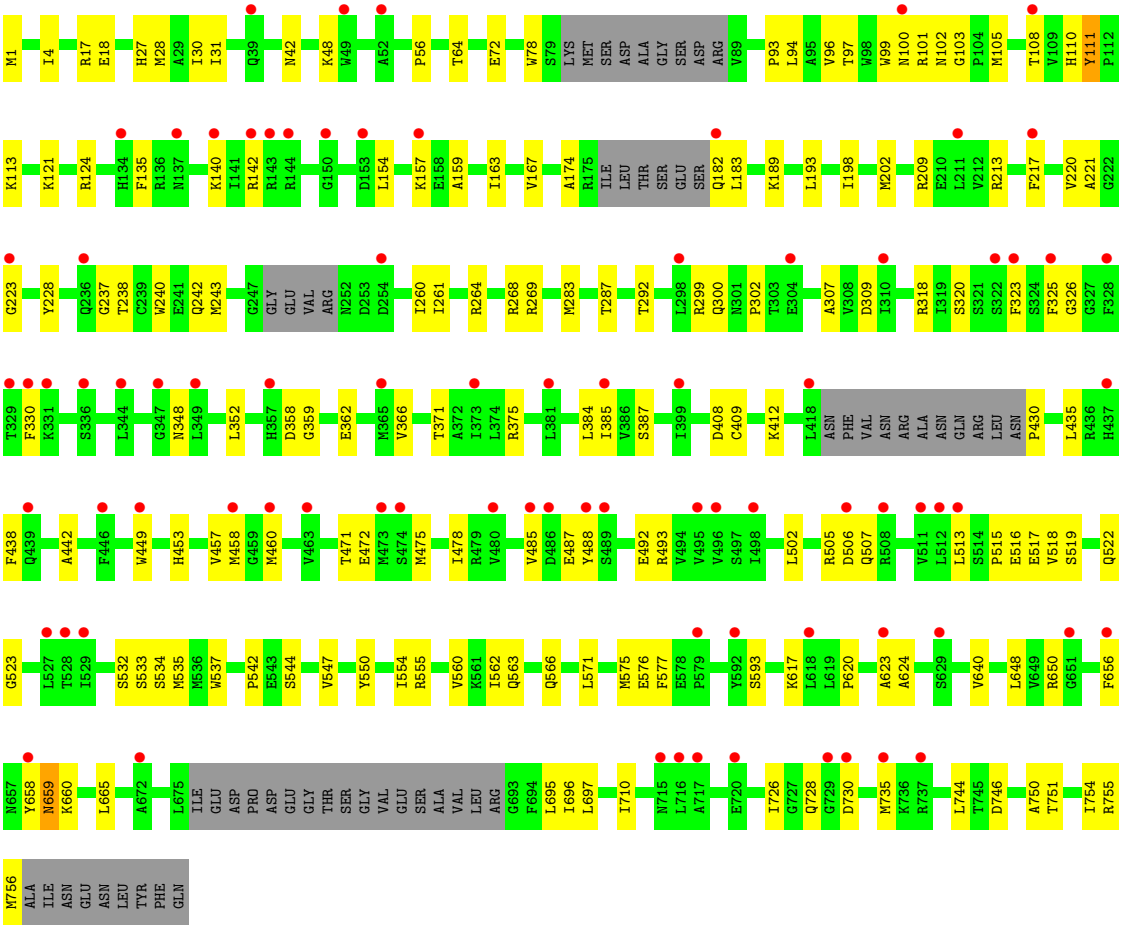






• Molecule 3: Polymerase basic protein 2

Chain L:


M1 I4 R17 E18 H27 M28 A29 I30 I31 Q39 N42 K48 W49 A52 P56 T64 E72 W78 S79 LYS MET SER ASP ALA GLY SER ASP ARG V89 P93 L94 A95 V96 T97 W98 W99 N100 R101 N102 G103 P104 M105 T108 V109 H110 Y111 P112
K113 K121 R124 H134 F135 R136 N137 K140 I141 R142 R143 R144 G150 D153 L154 K157 E158 A159 I163 V167 A174 R175 ILE LEU THR SER GLU SER Q182 L183 K189 L193 I198 M202 R209 E210 L211 V212 R213 F217 V220 A221 G222
G223 Y228 Q236 G237 T238 C239 W240 E241 Q242 M243 G247 GLY GLU VAL ARG N252 D253 D254 I260 I261 R264 R268 R269 M283 T287 T292 L298 R299 Q300 N301 P302 T303 E304 A307 V308 D309 I310 R318 I319 S320 S321 S322 F323 S324 F325 G326 G327 F328
T329 F330 K331 S336 L344 G347 N348 L349 L352 H357 D358 G359 E362 M365 V366 T371 A372 I373 L374 R375 L381 L384 I385 V386 S387 I399 D408 C409 K412 L418 ASN PHE VAL ASN ARG ALA ASN GLN ARG LEU ASN P430 L435 R436 H437
F438 Q439 A442 F446 W449 H453 V457 M458 G459 M460 V463 T471 E472 K473 S474 M475 I478 R479 V480 V485 D486 E487 Y488 S489 E492 R493 V494 V495 V496 S497 I498 L502 R505 D506 Q507 R508 V511 L512 L513 S514 P515 E516 E517 V518 S519 Q522
G523 L527 T528 I529 S532 S533 S534 M535 M536 W537 P542 E543 S544 V547 Y550 I554 R555 V560 K561 T562 Q563 Q566 L571 M575 E576 F577 E578 P579 Y592 S593 K617 L618 L619 P620 A623 A624 S629 V640 L648 V649 R650 G651 F656
N657 Y658 N659 K660 L665 A672 L675 ILE GLU ASP PRO ASP GLU GLY THR SER GLY VAL GLU SER ALA VAL LEU ARG G693 F694 L695 I696 L697 I710 W715 L716 A717 E720 I726 G727 Q728 G729 D730 M735 K736 R737 L744 T745 D746 A750 T751 I754 R755
M756 ALA ILE ASN GLU ASN LEU TYR PHE GLN

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	336.48Å 191.90Å 235.74Å 90.00° 91.50° 90.00°	Depositor
Resolution (Å)	138.62 – 3.31 138.62 – 3.31	Depositor EDS
% Data completeness (in resolution range)	55.1 (138.62-3.31) 55.1 (138.62-3.31)	Depositor EDS
R_{merge}	0.46	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.46 (at 3.33Å)	Xtriage
Refinement program	PHENIX 1.14_3260	Depositor
R, R_{free}	0.235 , 0.279 0.235 , 0.279	Depositor DCC
R_{free} test set	6172 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	88.5	Xtriage
Anisotropy	0.050	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 76.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.27$	Xtriage
Estimated twinning fraction	0.025 for -1/2*h-3/2*k,-1/2*h+1/2*k,-l 0.024 for -1/2*h+3/2*k,1/2*h+1/2*k,-l 0.027 for 1/2*h-3/2*k,-1/2*h-1/2*k,-l 0.024 for 1/2*h+3/2*k,1/2*h-1/2*k,-l 0.029 for -h,-k,l	Xtriage
F_o,F_c correlation	0.87	EDS
Total number of atoms	67232	wwPDB-VP
Average B, all atoms (Å^2)	117.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.07% 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

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	0.25	0/5853	0.46	0/7886
1	D	0.25	0/5720	0.44	0/7701
1	G	0.25	0/5827	0.45	0/7850
1	J	0.25	0/5833	0.45	0/7860
2	B	0.25	0/5529	0.45	1/7456 (0.0%)
2	E	0.25	0/5417	0.43	0/7306
2	H	0.25	0/5513	0.44	0/7433
2	K	0.25	0/5485	0.44	0/7396
3	C	0.24	0/5866	0.47	0/7900
3	F	0.24	0/5828	0.47	1/7851 (0.0%)
3	I	0.24	0/5883	0.46	0/7923
3	L	0.24	0/5738	0.46	0/7725
All	All	0.25	0/68492	0.45	2/92287 (0.0%)

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	A	0	1
3	C	0	1
All	All	0	2

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	579	LEU	CB-CG-CD2	-9.30	95.18	111.00
3	F	716	LEU	CA-CB-CG	6.75	130.82	115.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	297	HIS	Peptide
3	C	739	ARG	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	A	5733	0	5662	123	0
1	D	5604	0	5527	120	0
1	G	5707	0	5624	134	0
1	J	5713	0	5630	116	0
2	B	5426	0	5374	109	0
2	E	5315	0	5262	129	0
2	H	5410	0	5366	138	0
2	K	5382	0	5330	136	0
3	C	5772	0	5915	141	0
3	F	5734	0	5869	138	0
3	I	5789	0	5934	152	0
3	L	5647	0	5786	120	0
All	All	67232	0	67279	1334	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 10.

The worst 5 of 1334 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)
3:F:317:LEU:HD12	3:F:496:VAL:H	1.41	0.86
3:F:289:ILE:HG22	3:F:529:ILE:HG13	1.63	0.81
2:E:708:PRO:HA	3:F:728:GLN:HE21	1.46	0.79
3:F:149:PRO:HD2	3:F:209:ARG:HH12	1.49	0.77
2:B:309:TRP:HZ2	2:B:416:SER:HB3	1.49	0.77

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	696/716 (97%)	646 (93%)	49 (7%)	1 (0%)	51	81
1	D	674/716 (94%)	632 (94%)	42 (6%)	0	100	100
1	G	693/716 (97%)	644 (93%)	49 (7%)	0	100	100
1	J	694/716 (97%)	646 (93%)	48 (7%)	0	100	100
2	B	667/757 (88%)	628 (94%)	38 (6%)	1 (0%)	51	81
2	E	653/757 (86%)	621 (95%)	32 (5%)	0	100	100
2	H	664/757 (88%)	629 (95%)	34 (5%)	1 (0%)	47	76
2	K	661/757 (87%)	623 (94%)	37 (6%)	1 (0%)	47	76
3	C	716/765 (94%)	658 (92%)	56 (8%)	2 (0%)	41	71
3	F	712/765 (93%)	660 (93%)	52 (7%)	0	100	100
3	I	718/765 (94%)	658 (92%)	59 (8%)	1 (0%)	51	81
3	L	697/765 (91%)	647 (93%)	49 (7%)	1 (0%)	51	81
All	All	8245/8952 (92%)	7692 (93%)	545 (7%)	8 (0%)	51	81

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	298	GLU
2	B	295	ASP
2	K	429	LYS
3	I	745	THR
3	L	516	GLU

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	634/644 (98%)	623 (98%)	11 (2%)	60	79
1	D	620/644 (96%)	608 (98%)	12 (2%)	57	78
1	G	631/644 (98%)	622 (99%)	9 (1%)	67	82
1	J	632/644 (98%)	620 (98%)	12 (2%)	57	78
2	B	602/669 (90%)	593 (98%)	9 (2%)	65	81
2	E	589/669 (88%)	576 (98%)	13 (2%)	52	76
2	H	600/669 (90%)	590 (98%)	10 (2%)	60	79
2	K	597/669 (89%)	588 (98%)	9 (2%)	65	81
3	C	641/676 (95%)	631 (98%)	10 (2%)	62	80
3	F	637/676 (94%)	619 (97%)	18 (3%)	43	71
3	I	643/676 (95%)	635 (99%)	8 (1%)	71	84
3	L	628/676 (93%)	617 (98%)	11 (2%)	59	79
All	All	7454/7956 (94%)	7322 (98%)	132 (2%)	59	79

5 of 132 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	K	430	ARG
2	K	706	ARG
3	L	695	LEU
2	E	598	LEU
2	E	579	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 21 such sidechains are listed below:

Mol	Chain	Res	Type
3	I	614	GLN
1	J	217	GLN
3	L	728	GLN
2	K	292	ASN
1	J	146	HIS

5.3.3 RNA

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

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	704/716 (98%)	0.71	20 (2%) 53 51	36, 86, 152, 183	0
1	D	686/716 (95%)	1.11	121 (17%) 1 1	89, 142, 214, 261	0
1	G	701/716 (97%)	0.59	13 (1%) 66 65	39, 86, 147, 214	0
1	J	702/716 (98%)	0.87	75 (10%) 6 5	67, 121, 184, 240	0
2	B	679/757 (89%)	0.77	29 (4%) 35 35	35, 82, 137, 210	0
2	E	665/757 (87%)	1.40	177 (26%) 0 0	93, 164, 215, 239	0
2	H	676/757 (89%)	0.73	29 (4%) 35 35	35, 89, 157, 214	0
2	K	673/757 (88%)	0.92	81 (12%) 4 3	65, 124, 184, 221	0
3	C	724/765 (94%)	0.89	80 (11%) 5 5	46, 110, 167, 217	0
3	F	720/765 (94%)	0.98	116 (16%) 1 1	55, 134, 185, 220	0
3	I	726/765 (94%)	0.70	46 (6%) 20 21	42, 107, 163, 200	0
3	L	709/765 (92%)	0.89	83 (11%) 4 3	70, 123, 173, 215	0
All	All	8365/8952 (93%)	0.88	870 (10%) 6 6	35, 114, 186, 261	0

The worst 5 of 870 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	E	440	GLY	12.8
3	L	717	ALA	11.4
3	C	250	VAL	9.8
3	F	140	LYS	9.7
2	E	439	ASP	8.3

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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.