wwPDB X-ray Structure Validation Summary Report (i)

Sep 19, 2023 – 09:11 PM EDT

PDB ID	:	5IKN
Title	:	Crystal Structure of the T7 Replisome in the Absence of DNA
Authors	:	Wallen, J.R.; Ellenberger, T.
Deposited on	:	2016-03-03
Resolution	:	4.80 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:

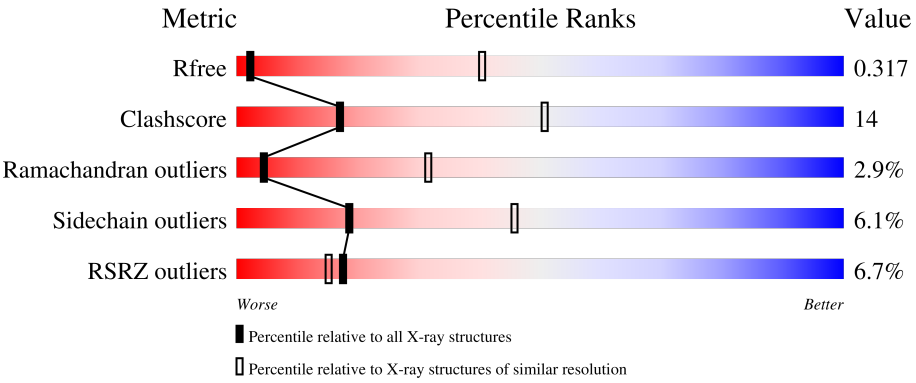
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 4.80 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	1096 (5.80-3.80)
Clashscore	141614	1170 (5.80-3.80)
Ramachandran outliers	138981	1105 (5.80-3.80)
Sidechain outliers	138945	1085 (5.80-3.80)
RSRZ outliers	127900	1126 (5.90-3.70)

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	704	4% 65% 22% · 9%
1	B	704	2% 65% 23% · 10%
1	C	704	5% 65% 23% · 10%
2	D	486	10% 66% 30% ·
2	E	486	4% 64% 32% ·

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	486	8% 61% 34% 5%
2	G	486	9% 63% 31%
2	H	486	6% 67% 29%
2	I	486	8% 55% 34% 7%
2	J	486	5% 64% 31%
3	K	105	21% 79% 18%
3	L	105	3% 67% 30%
3	M	105	25% 85% 14%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 43846 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 DNA-directed DNA polymerase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	639	Total	C	N	O	S	0	0	0
			5092	3236	888	947	21			
1	B	636	Total	C	N	O	S	0	0	0
			5073	3226	884	942	21			
1	C	633	Total	C	N	O	S	0	0	0
			5041	3201	880	939	21			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	5	ALA	ASP	engineered mutation	UNP P00581
A	7	ALA	GLU	engineered mutation	UNP P00581
B	5	ALA	ASP	engineered mutation	UNP P00581
B	7	ALA	GLU	engineered mutation	UNP P00581
C	5	ALA	ASP	engineered mutation	UNP P00581
C	7	ALA	GLU	engineered mutation	UNP P00581

- Molecule 2 is a protein called DNA primase/helicase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	485	Total	C	N	O	S	0	0	0
			3763	2353	660	725	25			
2	E	485	Total	C	N	O	S	0	0	0
			3763	2353	660	725	25			
2	F	486	Total	C	N	O	S	0	0	0
			3768	2355	661	727	25			
2	G	484	Total	C	N	O	S	0	0	0
			3757	2350	659	723	25			
2	H	486	Total	C	N	O	S	0	0	0
			3768	2355	661	727	25			
2	I	472	Total	C	N	O	S	0	0	0
			3670	2295	647	703	25			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	J	483	Total	C	N	O	S	0	0	0
			3745	2341	658	721	25			

- Molecule 3 is a protein called Thioredoxin-1.

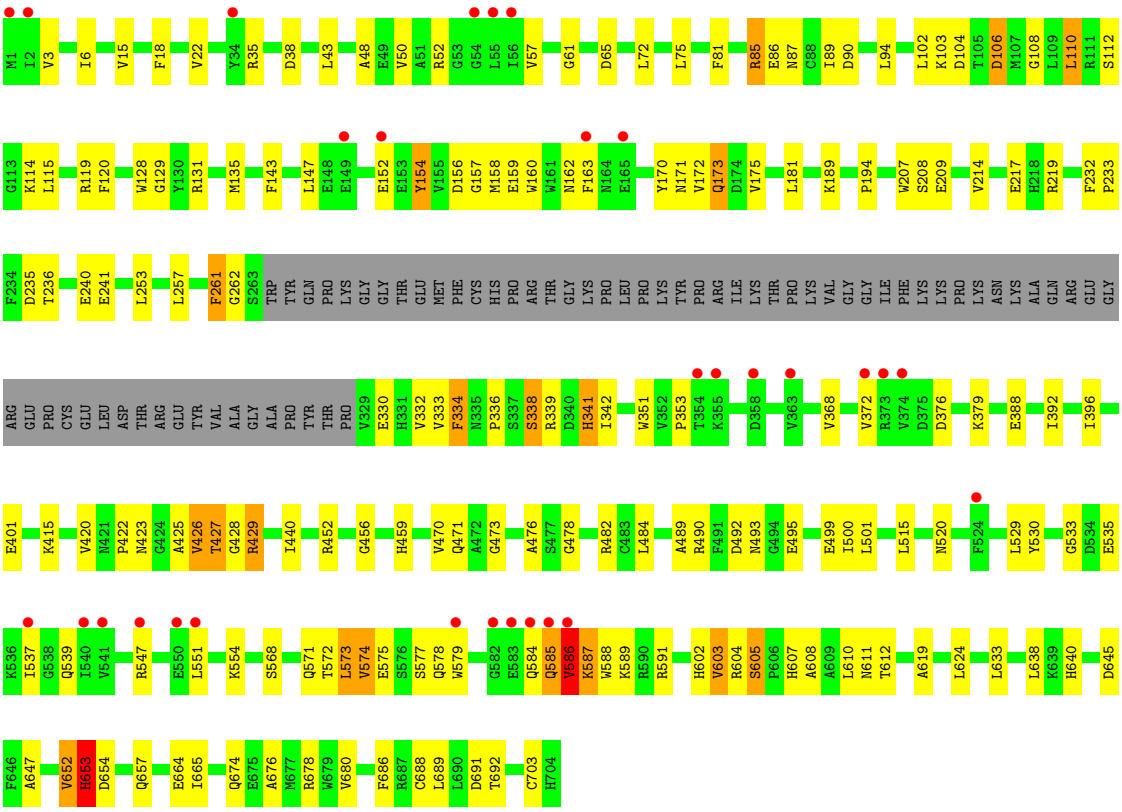
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	K	105	Total	C	N	O	S	0	0	0
			802	518	129	152	3			
3	L	105	Total	C	N	O	S	0	0	0
			802	518	129	152	3			
3	M	105	Total	C	N	O	S	0	0	0
			802	518	129	152	3			

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.

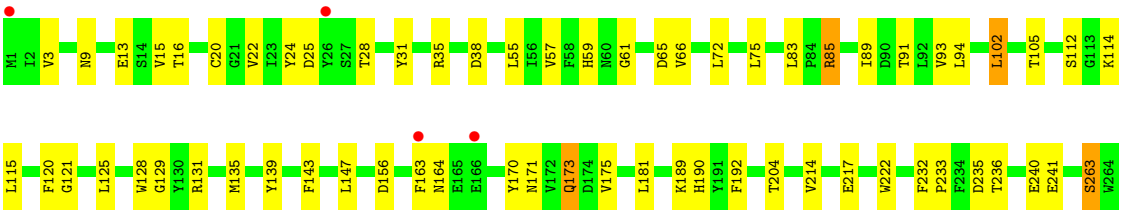
• Molecule 1: DNA-directed DNA polymerase

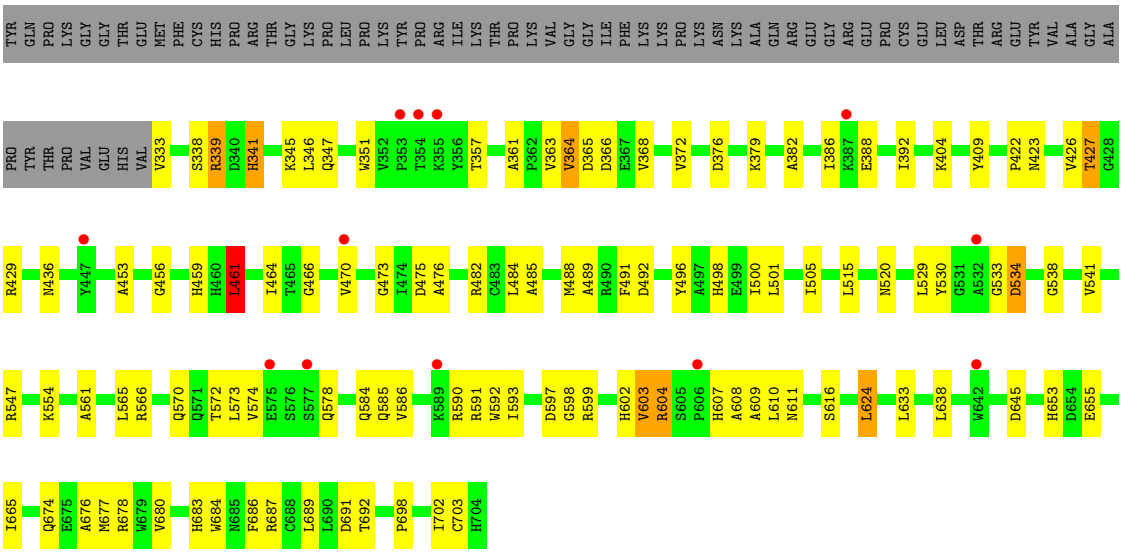
Chain A:



• Molecule 1: DNA-directed DNA polymerase

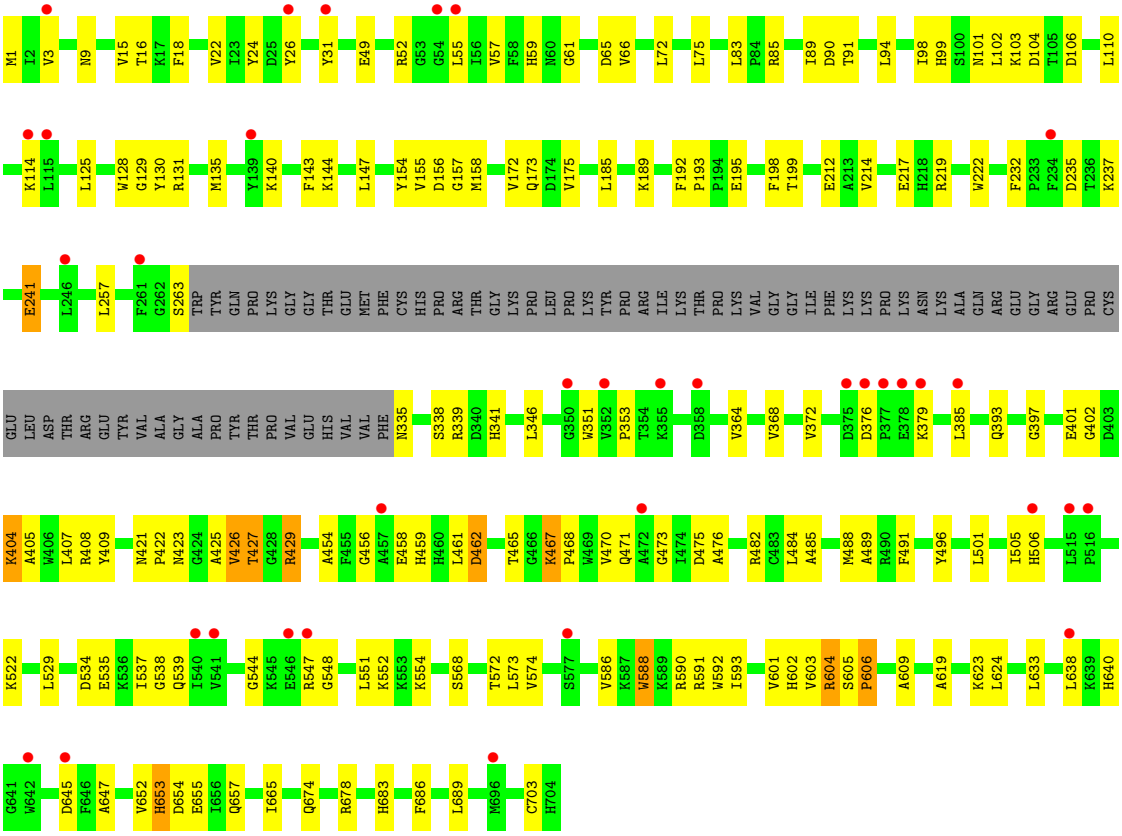
Chain B:





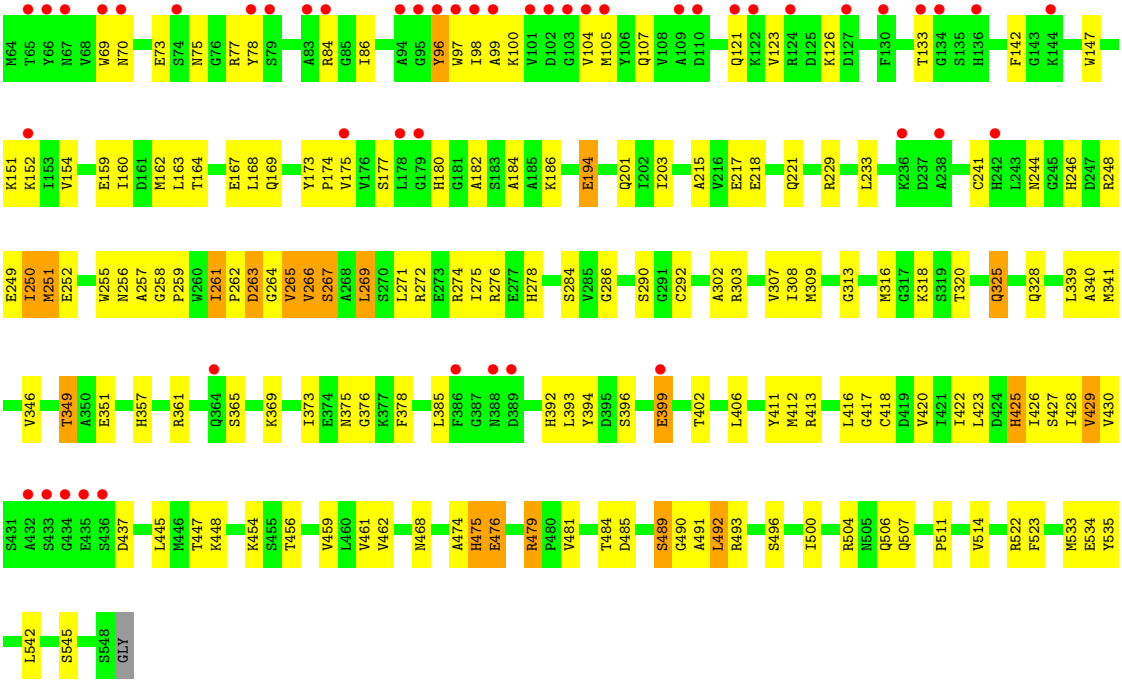
• Molecule 1: DNA-directed DNA polymerase

Chain C:



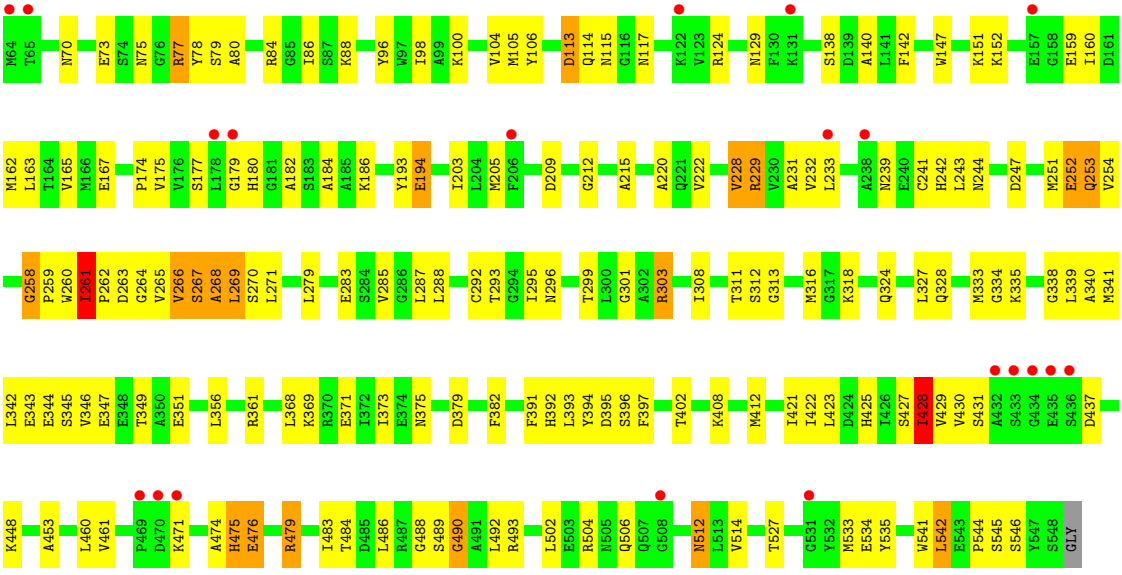
• Molecule 2: DNA primase/helicase

Chain D:



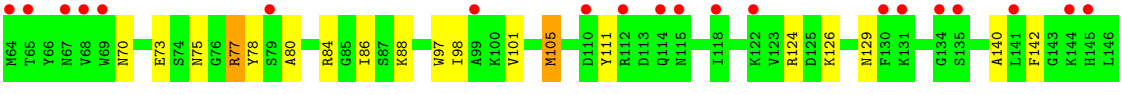
• Molecule 2: DNA primase/helicase

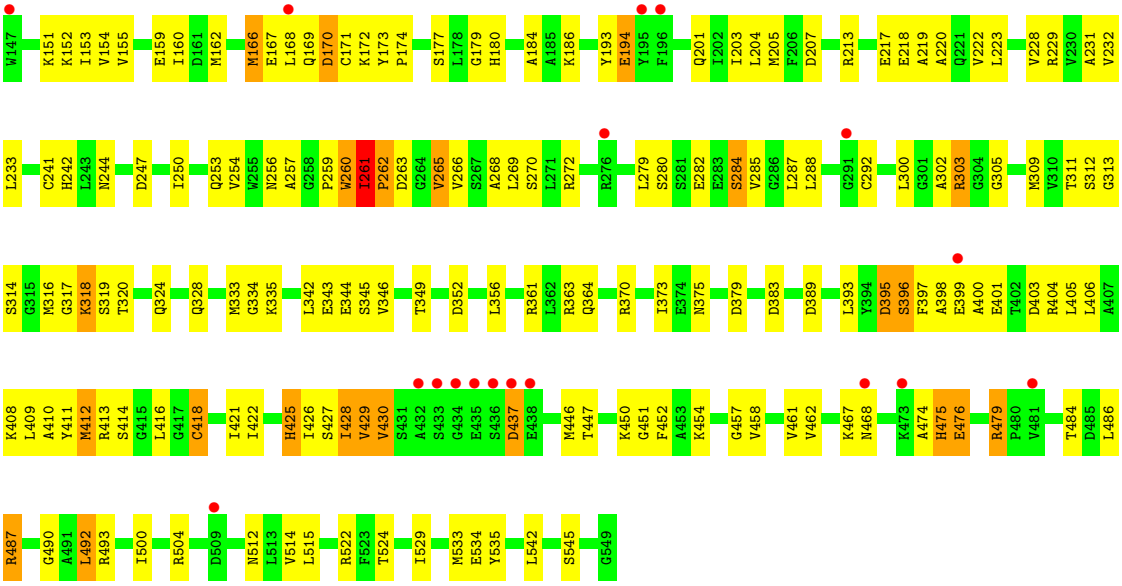
Chain E:



• Molecule 2: DNA primase/helicase

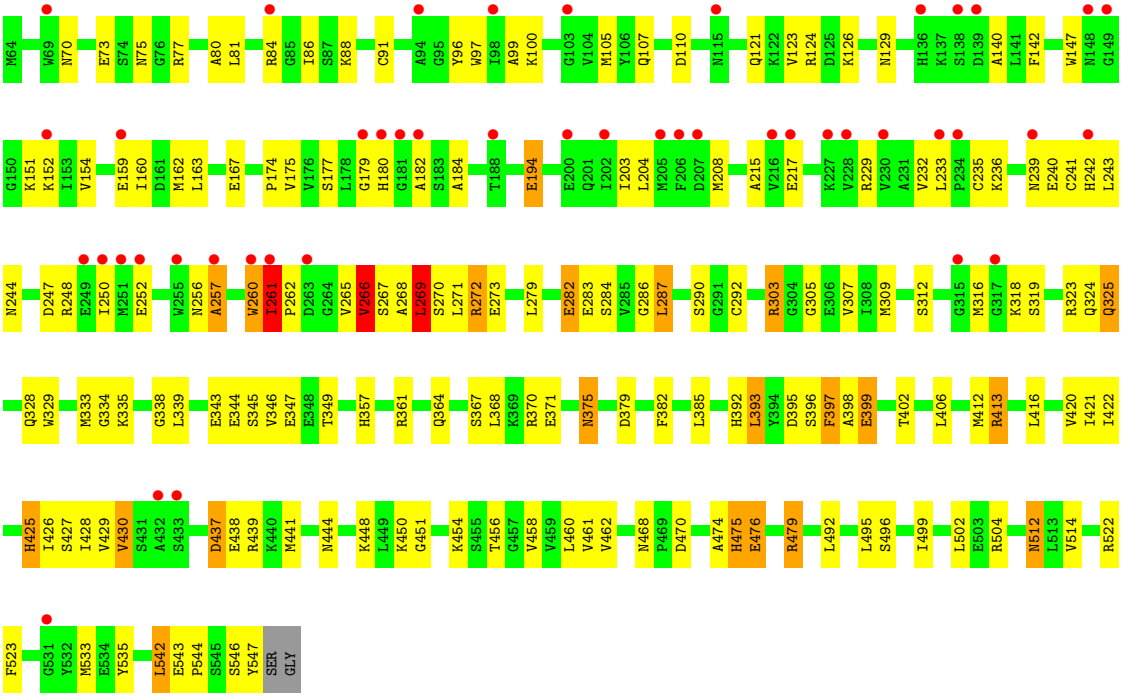
Chain F:





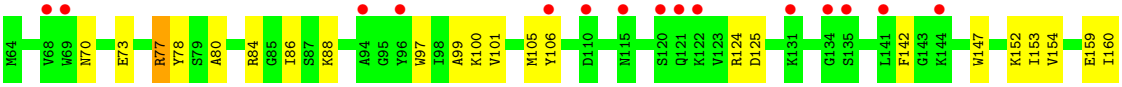
• Molecule 2: DNA primase/helicase

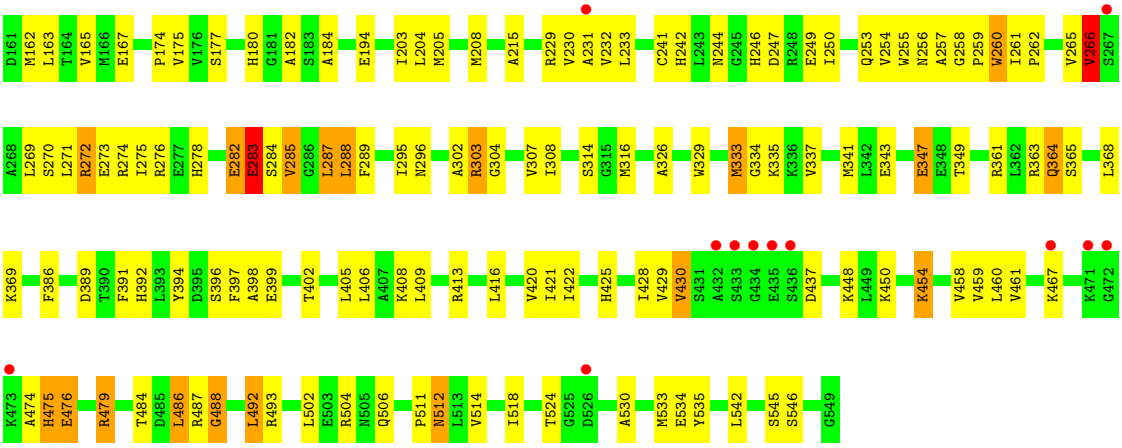
Chain G:



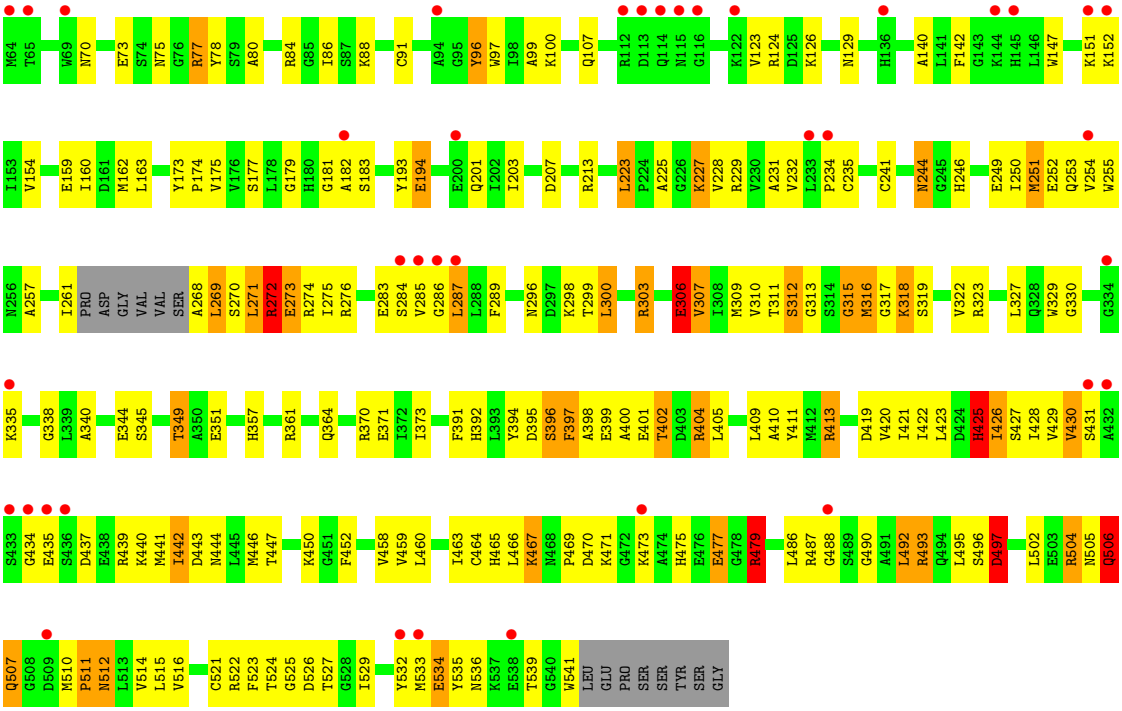
• Molecule 2: DNA primase/helicase

Chain H:

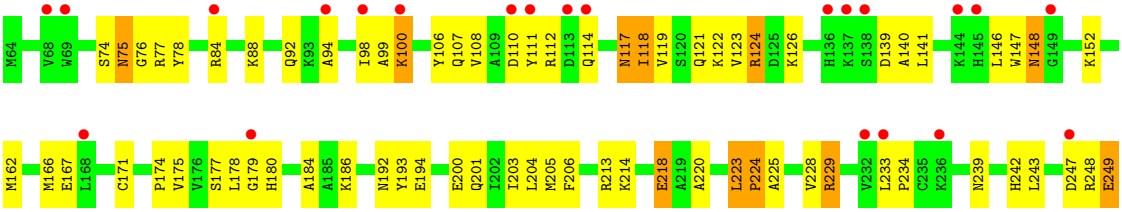


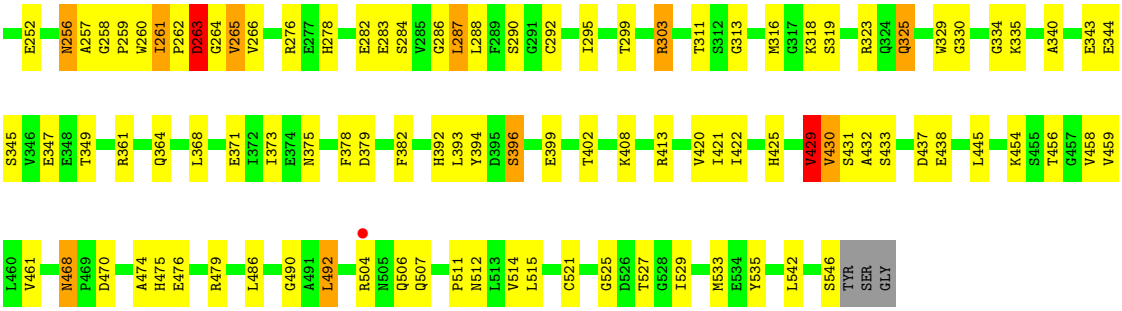


• Molecule 2: DNA primase/helicase

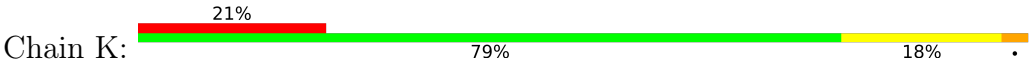


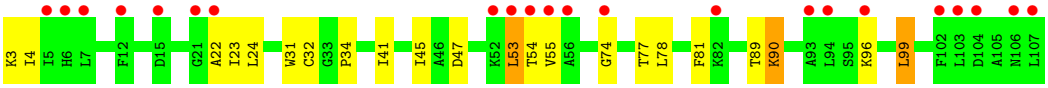
• Molecule 2: DNA primase/helicase



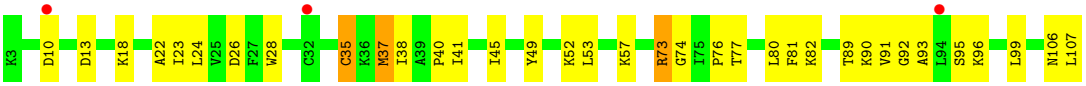


• Molecule 3: Thioredoxin-1

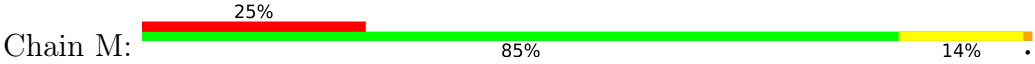


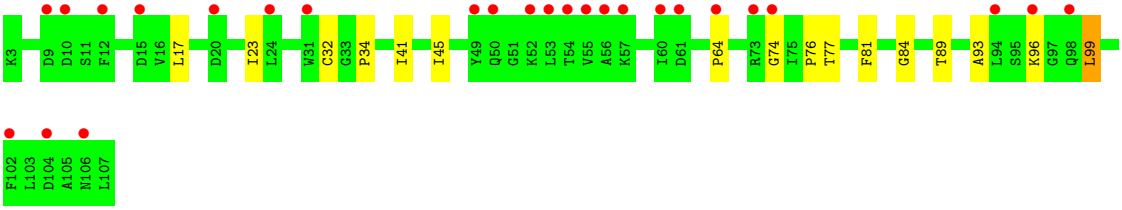


• Molecule 3: Thioredoxin-1



• Molecule 3: Thioredoxin-1





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	174.87Å 238.09Å 243.47Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.98 – 4.80 49.20 – 4.80	Depositor EDS
% Data completeness (in resolution range)	82.2 (29.98-4.80) 82.4 (49.20-4.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.05 (at 4.86Å)	Xtriage
Refinement program	PHENIX 1.10_2155	Depositor
R, R_{free}	0.275 , 0.318 0.274 , 0.317	Depositor DCC
R_{free} test set	2113 reflections (5.10%)	wwPDB-VP
Wilson B-factor (Å^2)	177.0	Xtriage
Anisotropy	0.290	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 201.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.27$	Xtriage
Estimated twinning fraction	0.006 for -h,l,k	Xtriage
F_o,F_c correlation	0.86	EDS
Total number of atoms	43846	wwPDB-VP
Average B, all atoms (Å^2)	240.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.28% 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	0.26	0/5211	0.42	0/7045
1	B	0.26	0/5193	0.41	0/7021
1	C	0.25	0/5158	0.40	0/6972
2	D	0.26	0/3826	0.44	0/5152
2	E	0.28	0/3826	0.48	0/5152
2	F	0.28	0/3831	0.49	0/5157
2	G	0.26	0/3820	0.45	0/5144
2	H	0.27	0/3831	0.48	0/5157
2	I	0.35	2/3729 (0.1%)	0.55	4/5016 (0.1%)
2	J	0.26	0/3807	0.46	0/5126
3	K	0.25	0/817	0.43	0/1108
3	L	0.27	0/817	0.46	0/1108
3	M	0.25	0/817	0.43	0/1108
All	All	0.27	2/44683 (0.0%)	0.46	4/60266 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	I	312	SER	CA-CB	7.75	1.64	1.52
2	I	318	LYS	CD-CE	5.98	1.66	1.51

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	I	318	LYS	CD-CE-NZ	13.48	142.71	111.70
2	I	312	SER	N-CA-CB	7.40	121.60	110.50
2	I	318	LYS	CA-CB-CG	-5.66	100.96	113.40
2	I	312	SER	CB-CA-C	-5.31	100.02	110.10

There are no chirality outliers.

There are no planarity outliers.

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	5092	0	4986	118	0
1	B	5073	0	4965	102	0
1	C	5041	0	4937	100	0
2	D	3763	0	3724	123	0
2	E	3763	0	3724	125	0
2	F	3768	0	3727	147	1
2	G	3757	0	3719	134	0
2	H	3768	0	3727	131	0
2	I	3670	0	3638	184	0
2	J	3745	0	3710	123	0
3	K	802	0	816	15	0
3	L	802	0	816	21	1
3	M	802	0	816	10	0
All	All	43846	0	43305	1221	1

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 14.

The worst 5 of 1221 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:I:318:LYS:NZ	2:I:502:LEU:H	1.20	1.34
2:I:318:LYS:HZ1	2:I:502:LEU:N	1.34	1.25
2:I:318:LYS:HE3	2:I:502:LEU:HB2	1.47	0.95
2:I:312:SER:HB3	2:I:318:LYS:CE	1.97	0.94
2:I:318:LYS:NZ	2:I:502:LEU:N	2.03	0.93

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:F:75:ASN:O	3:L:73:ARG:NH2[2_455]	2.17	0.03

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	635/704 (90%)	579 (91%)	44 (7%)	12 (2%)	8 40
1	B	632/704 (90%)	577 (91%)	44 (7%)	11 (2%)	9 43
1	C	629/704 (89%)	568 (90%)	50 (8%)	11 (2%)	9 43
2	D	483/486 (99%)	418 (86%)	52 (11%)	13 (3%)	5 33
2	E	483/486 (99%)	414 (86%)	51 (11%)	18 (4%)	3 27
2	F	484/486 (100%)	403 (83%)	60 (12%)	21 (4%)	2 24
2	G	482/486 (99%)	411 (85%)	52 (11%)	19 (4%)	3 26
2	H	484/486 (100%)	409 (84%)	56 (12%)	19 (4%)	3 26
2	I	468/486 (96%)	382 (82%)	63 (14%)	23 (5%)	2 22
2	J	481/486 (99%)	413 (86%)	51 (11%)	17 (4%)	3 28
3	K	103/105 (98%)	97 (94%)	6 (6%)	0	100 100
3	L	103/105 (98%)	93 (90%)	10 (10%)	0	100 100
3	M	103/105 (98%)	98 (95%)	5 (5%)	0	100 100
All	All	5570/5829 (96%)	4862 (87%)	544 (10%)	164 (3%)	4 31

5 of 164 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	422	PRO
1	A	426	VAL
1	A	586	VAL
1	B	346	LEU
1	C	156	ASP

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	527/582 (90%)	492 (93%)	35 (7%)	16	43
1	B	524/582 (90%)	502 (96%)	22 (4%)	30	55
1	C	521/582 (90%)	505 (97%)	16 (3%)	40	62
2	D	403/403 (100%)	381 (94%)	22 (6%)	21	49
2	E	403/403 (100%)	372 (92%)	31 (8%)	13	39
2	F	403/403 (100%)	379 (94%)	24 (6%)	19	46
2	G	402/403 (100%)	381 (95%)	21 (5%)	23	49
2	H	403/403 (100%)	383 (95%)	20 (5%)	24	50
2	I	391/403 (97%)	344 (88%)	47 (12%)	5	22
2	J	401/403 (100%)	369 (92%)	32 (8%)	12	37
3	K	85/85 (100%)	81 (95%)	4 (5%)	26	52
3	L	85/85 (100%)	79 (93%)	6 (7%)	14	41
3	M	85/85 (100%)	83 (98%)	2 (2%)	49	69
All	All	4633/4822 (96%)	4351 (94%)	282 (6%)	18	46

5 of 282 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	I	507	GLN
2	J	100	LYS
2	J	368	LEU
2	E	113	ASP
2	D	507	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
2	F	328	GLN
2	F	375	ASN
2	J	192	ASN
2	H	425	HIS
2	I	506	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

There are no ligands in this entry.

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, 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	639/704 (90%)	0.31	30 (4%) 31 27	137, 214, 280, 318	0
1	B	636/704 (90%)	0.26	16 (2%) 57 48	135, 192, 256, 283	0
1	C	633/704 (89%)	0.45	35 (5%) 25 22	180, 240, 332, 354	0
2	D	485/486 (99%)	0.56	49 (10%) 7 7	190, 241, 349, 382	0
2	E	485/486 (99%)	0.32	20 (4%) 37 31	174, 220, 262, 278	0
2	F	486/486 (100%)	0.49	38 (7%) 13 12	177, 230, 274, 298	0
2	G	484/486 (99%)	0.53	46 (9%) 8 8	170, 238, 358, 443	0
2	H	486/486 (100%)	0.46	27 (5%) 24 21	158, 232, 285, 346	0
2	I	472/486 (97%)	0.57	38 (8%) 12 11	207, 282, 332, 359	0
2	J	483/486 (99%)	0.33	23 (4%) 30 27	174, 244, 298, 378	0
3	K	105/105 (100%)	1.26	22 (20%) 1 1	219, 359, 416, 451	0
3	L	105/105 (100%)	0.53	3 (2%) 51 41	225, 255, 277, 292	0
3	M	105/105 (100%)	1.39	26 (24%) 0 1	242, 345, 382, 388	0
All	All	5604/5829 (96%)	0.46	373 (6%) 17 15	135, 233, 335, 451	0

The worst 5 of 373 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	H	472	GLY	7.7
3	K	22	ALA	7.1
1	A	584	GLN	6.4
2	D	102	ASP	5.9
2	I	433	SER	5.9

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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.