wwPDB X-ray Structure Validation Summary Report (i)

May 13, 2020 – 09:52 am BST

PDB ID	:	4M4W
Title	:	Mechanistic implications for the bacterial primosome assembly of the structure of a helicase-helicase loader complex
Authors	:	Liu, B.; Eliason, W.K.; Steitz, T.A.
Deposited on	:	2013-08-07
Resolution	:	6.10 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:

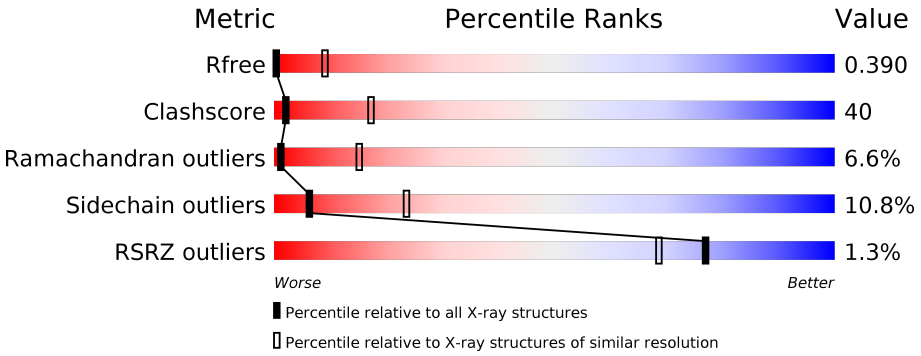
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 6.10 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.



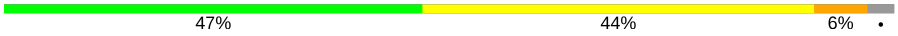
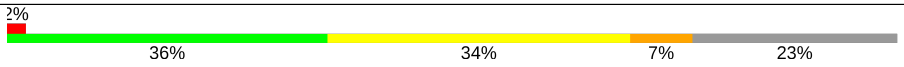
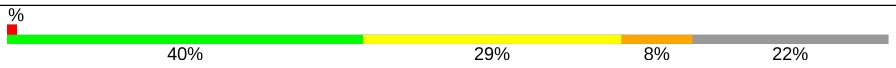
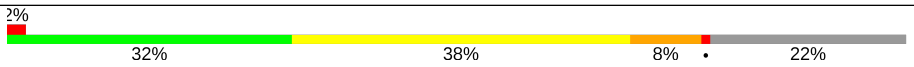
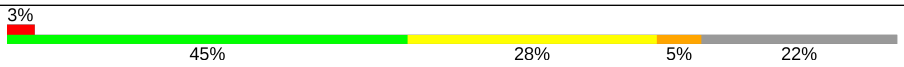
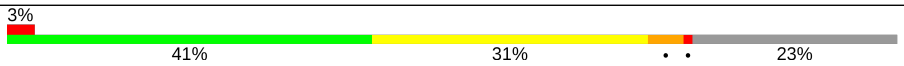
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1002 (8.30-3.88)
Clashscore	141614	1051 (8.30-3.90)
Ramachandran outliers	138981	1018 (8.30-3.86)
Sidechain outliers	138945	1019 (8.30-3.82)
RSRZ outliers	127900	1015 (8.20-3.78)

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 on 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	454	36% 35% 7% • 21%
1	B	454	41% 31% 6% 21%
1	C	454	33% 39% 6% • 21%
1	D	454	43% 30% 6% 21%
1	E	454	35% 35% 8% 21%
1	F	454	36% 34% 8% • 21%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	G	143	
2	H	143	
2	I	143	
3	J	317	
3	K	317	
3	L	317	
3	M	317	
3	N	317	
3	O	317	

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 31981 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 Replicative helicase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	358	Total	C	N	O	S	0	0	0
			2763	1739	473	538	13			
1	B	358	Total	C	N	O	S	0	0	0
			2763	1739	473	538	13			
1	C	358	Total	C	N	O	S	0	0	0
			2763	1739	473	538	13			
1	D	358	Total	C	N	O	S	0	0	0
			2763	1739	473	538	13			
1	E	358	Total	C	N	O	S	0	0	0
			2763	1739	473	538	13			
1	F	358	Total	C	N	O	S	0	0	0
			2763	1739	473	538	13			

- Molecule 2 is a protein called DNA primase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	G	138	Total	C	N	O	Se	0	0	0
			1122	712	199	205	6			
2	H	138	Total	C	N	O	Se	0	0	0
			1122	712	199	205	6			
2	I	138	Total	C	N	O	Se	0	0	0
			1122	712	199	205	6			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	530	GLU	ASP	CONFLICT	UNP Q9X4D0
G	531	LEU	VAL	CONFLICT	UNP Q9X4D0
H	530	GLU	ASP	CONFLICT	UNP Q9X4D0
H	531	LEU	VAL	CONFLICT	UNP Q9X4D0
I	530	GLU	ASP	CONFLICT	UNP Q9X4D0
I	531	LEU	VAL	CONFLICT	UNP Q9X4D0

- Molecule 3 is a protein called Primosomal protein DnaI.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	J	245	Total	C	N	O	S	0	0	0
			1992	1268	332	380	12			
3	K	248	Total	C	N	O	S	0	0	0
			2013	1285	329	387	12			
3	L	247	Total	C	N	O	S	0	0	0
			2010	1280	334	383	13			
3	M	247	Total	C	N	O	S	0	0	0
			2008	1281	327	387	13			
3	N	244	Total	C	N	O	S	0	0	0
			1983	1263	331	377	12			
3	O	250	Total	C	N	O	S	0	0	0
			2031	1297	331	390	13			

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
J	312	HIS	-	EXPRESSION TAG	UNP P06567
J	313	HIS	-	EXPRESSION TAG	UNP P06567
J	314	HIS	-	EXPRESSION TAG	UNP P06567
J	315	HIS	-	EXPRESSION TAG	UNP P06567
J	316	HIS	-	EXPRESSION TAG	UNP P06567
J	317	HIS	-	EXPRESSION TAG	UNP P06567
K	312	HIS	-	EXPRESSION TAG	UNP P06567
K	313	HIS	-	EXPRESSION TAG	UNP P06567
K	314	HIS	-	EXPRESSION TAG	UNP P06567
K	315	HIS	-	EXPRESSION TAG	UNP P06567
K	316	HIS	-	EXPRESSION TAG	UNP P06567
K	317	HIS	-	EXPRESSION TAG	UNP P06567
L	312	HIS	-	EXPRESSION TAG	UNP P06567
L	313	HIS	-	EXPRESSION TAG	UNP P06567
L	314	HIS	-	EXPRESSION TAG	UNP P06567
L	315	HIS	-	EXPRESSION TAG	UNP P06567
L	316	HIS	-	EXPRESSION TAG	UNP P06567
L	317	HIS	-	EXPRESSION TAG	UNP P06567
M	312	HIS	-	EXPRESSION TAG	UNP P06567
M	313	HIS	-	EXPRESSION TAG	UNP P06567
M	314	HIS	-	EXPRESSION TAG	UNP P06567
M	315	HIS	-	EXPRESSION TAG	UNP P06567
M	316	HIS	-	EXPRESSION TAG	UNP P06567
M	317	HIS	-	EXPRESSION TAG	UNP P06567
N	312	HIS	-	EXPRESSION TAG	UNP P06567
N	313	HIS	-	EXPRESSION TAG	UNP P06567

Continued on next page...

Continued from previous page...

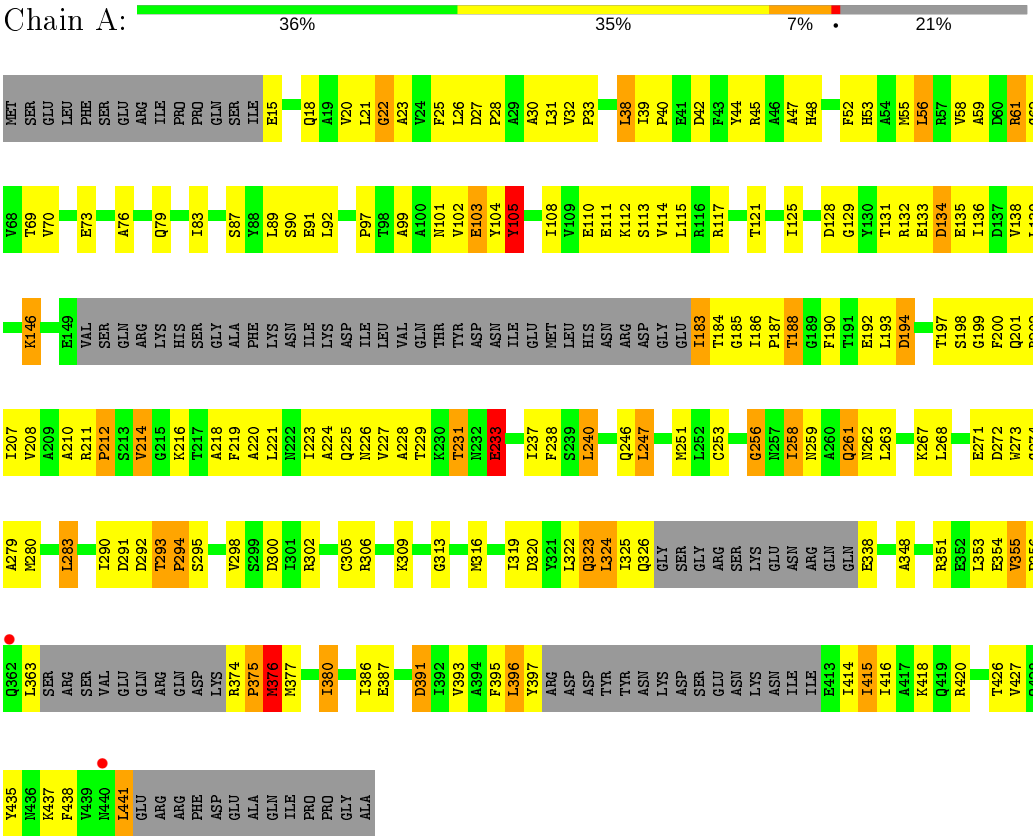
Chain	Residue	Modelled	Actual	Comment	Reference
N	314	HIS	-	EXPRESSION TAG	UNP P06567
N	315	HIS	-	EXPRESSION TAG	UNP P06567
N	316	HIS	-	EXPRESSION TAG	UNP P06567
N	317	HIS	-	EXPRESSION TAG	UNP P06567
O	312	HIS	-	EXPRESSION TAG	UNP P06567
O	313	HIS	-	EXPRESSION TAG	UNP P06567
O	314	HIS	-	EXPRESSION TAG	UNP P06567
O	315	HIS	-	EXPRESSION TAG	UNP P06567
O	316	HIS	-	EXPRESSION TAG	UNP P06567
O	317	HIS	-	EXPRESSION TAG	UNP P06567

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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.

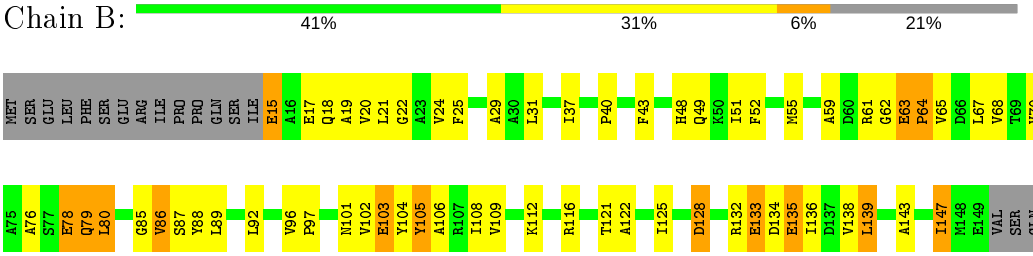
- Molecule 1: Replicative helicase

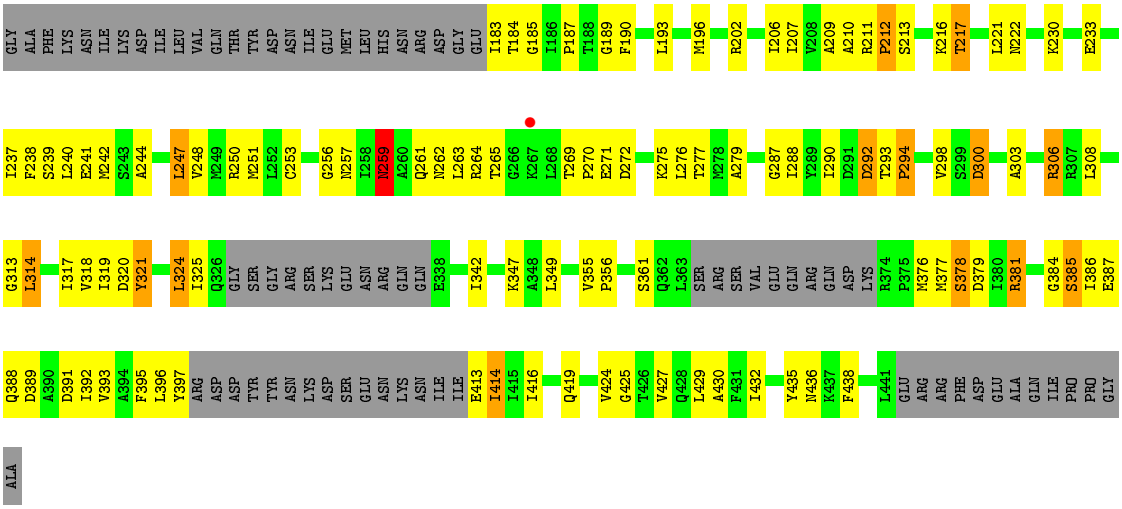
Chain A:



- Molecule 1: Replicative helicase

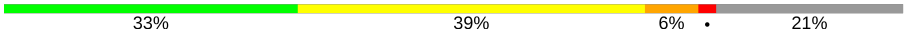
Chain B:

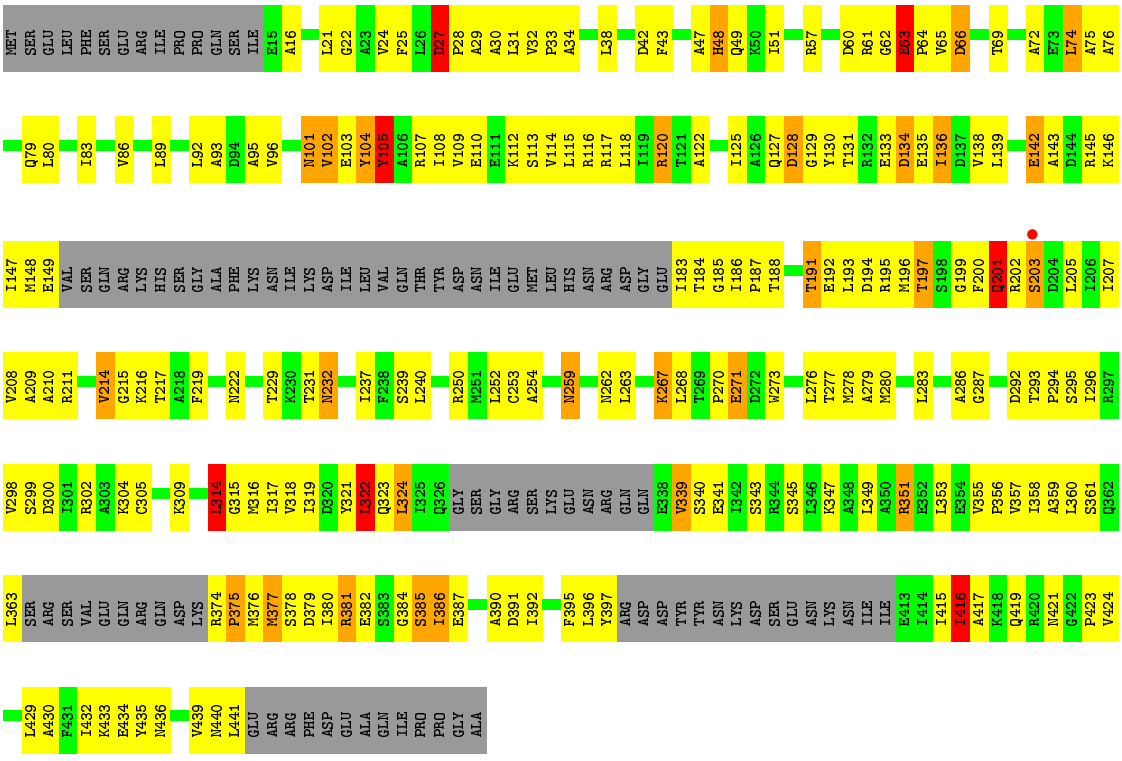




• Molecule 1: Replicative helicase

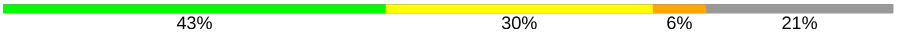
Chain C:

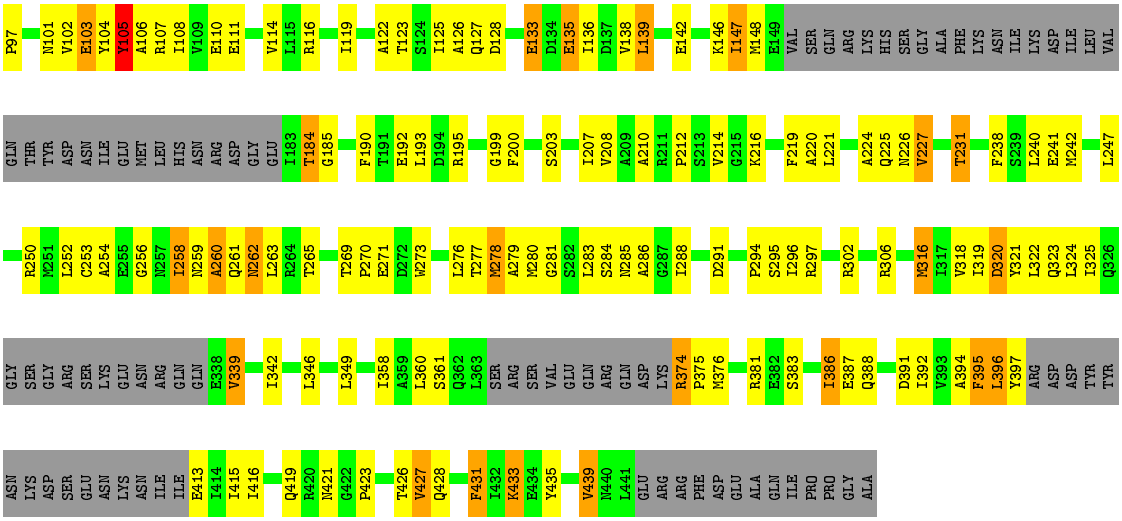




• Molecule 1: Replicative helicase

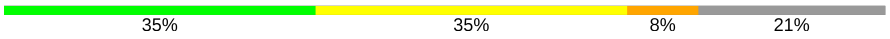
Chain D:

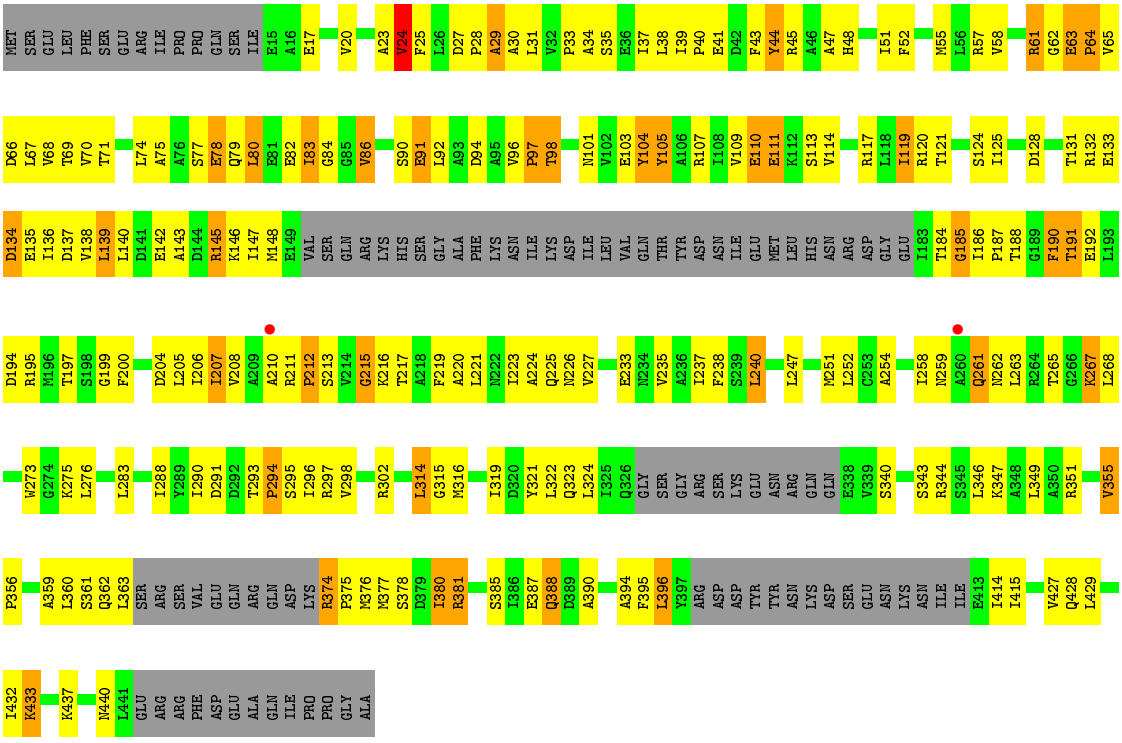




• Molecule 1: Replicative helicase

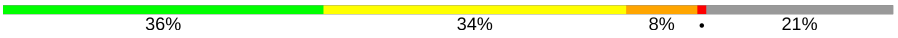
Chain E:

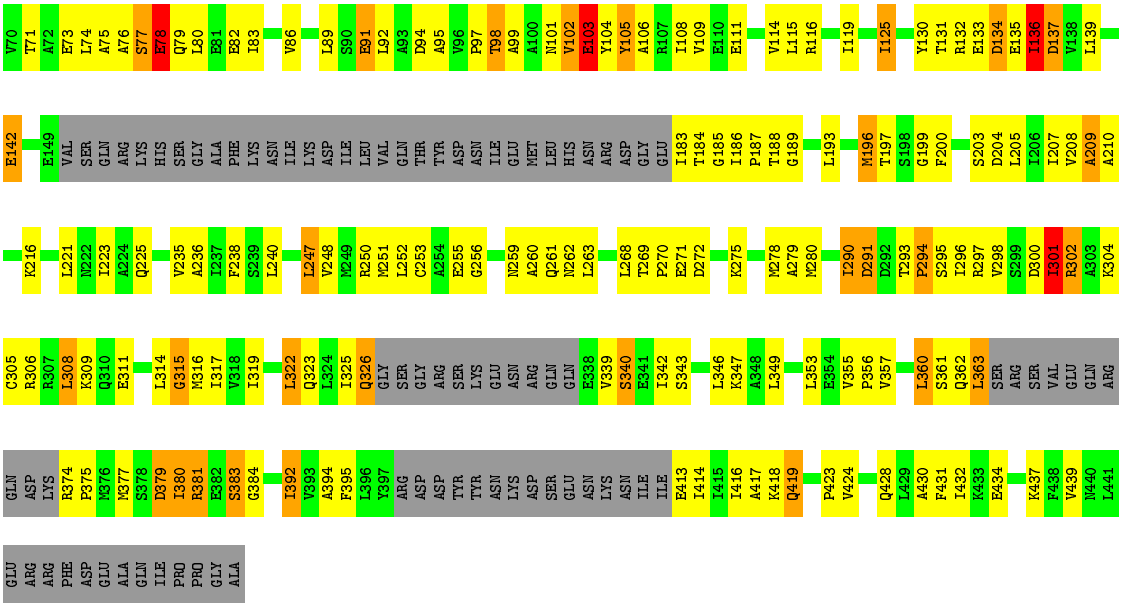




• Molecule 1: Replicative helicase

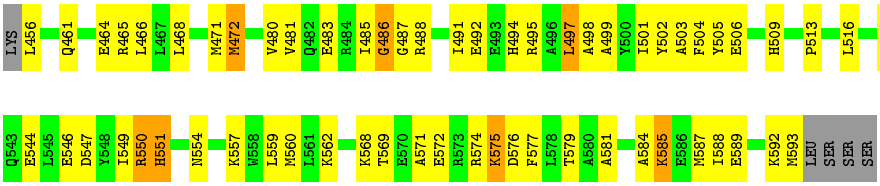
Chain F:





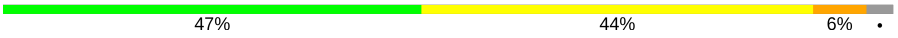
• Molecule 2: DNA primase

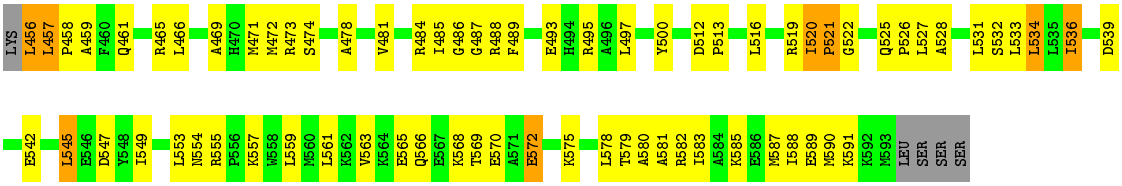
Chain G:



• Molecule 2: DNA primase

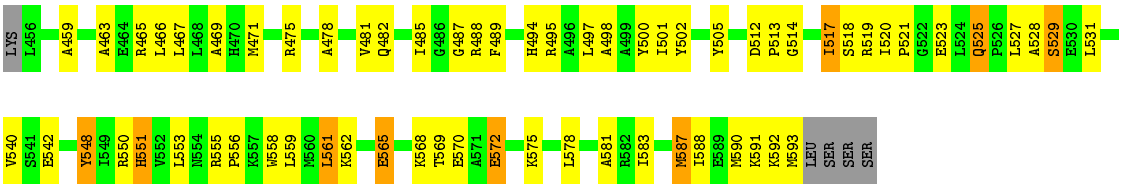
Chain H:





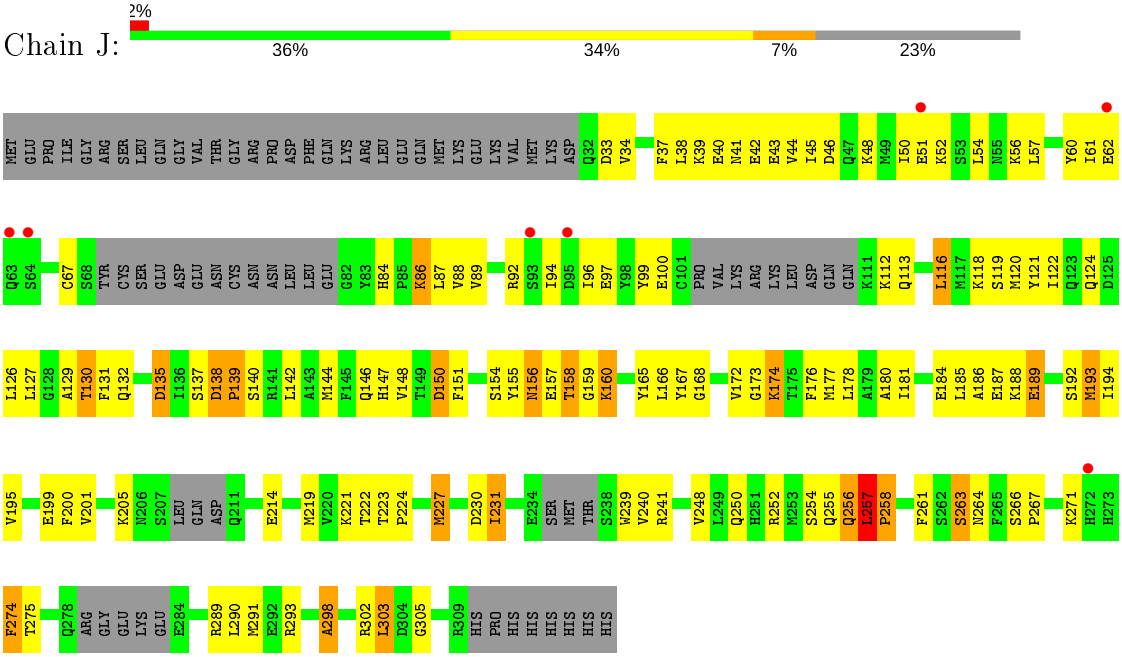
• Molecule 2: DNA primase

Chain I:



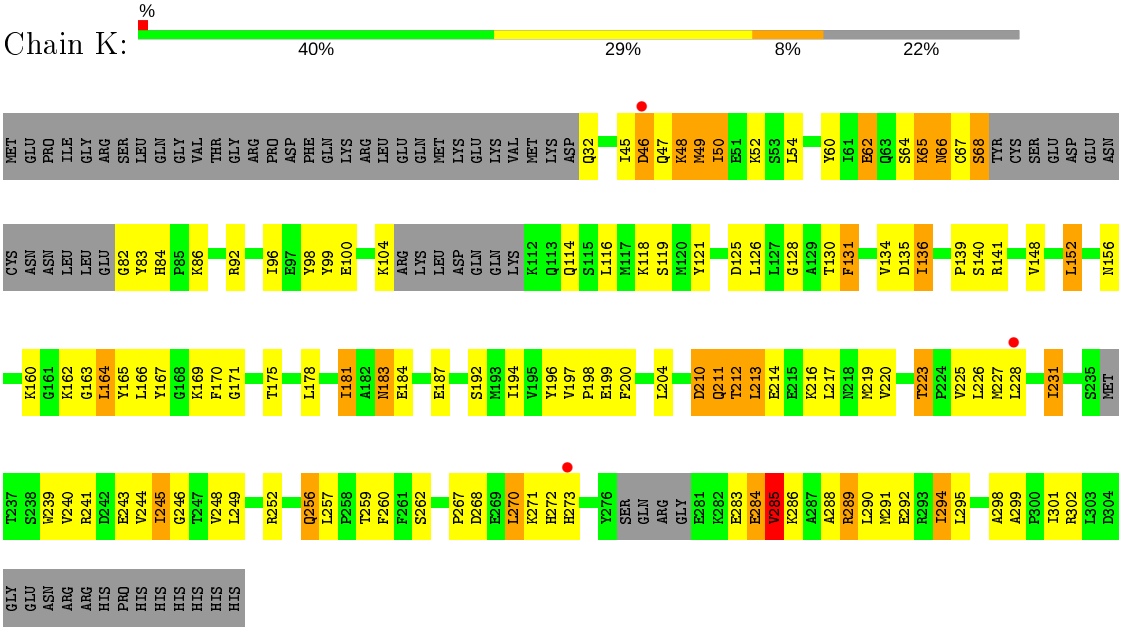
• Molecule 3: Primosomal protein DnaI

Chain J:



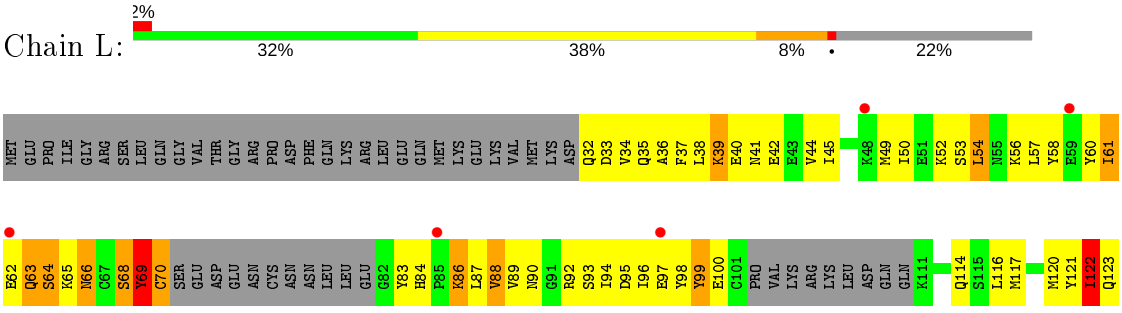
• Molecule 3: Primosomal protein DnaI

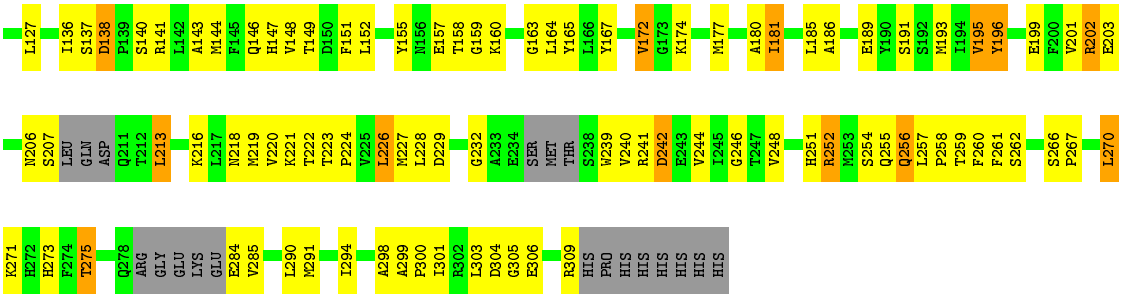
Chain K:



• Molecule 3: Primosomal protein DnaI

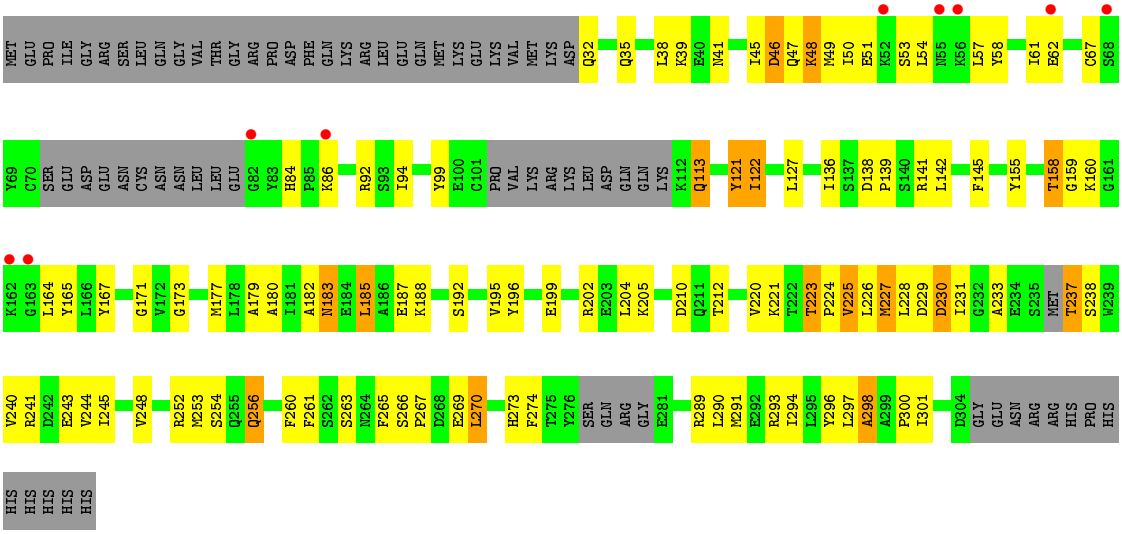
Chain L:





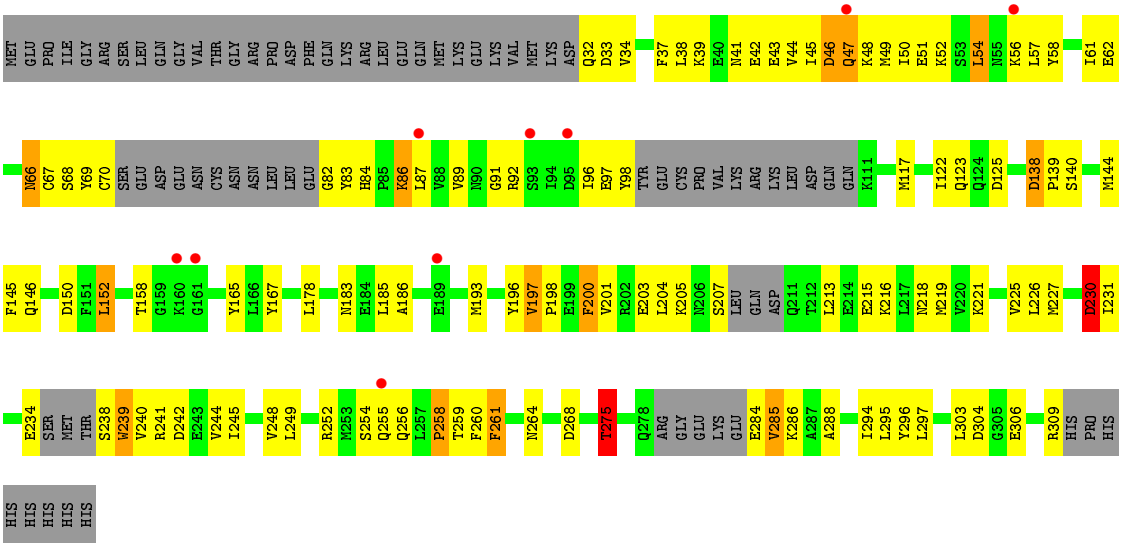
• Molecule 3: Primosomal protein DnaI

Chain M:



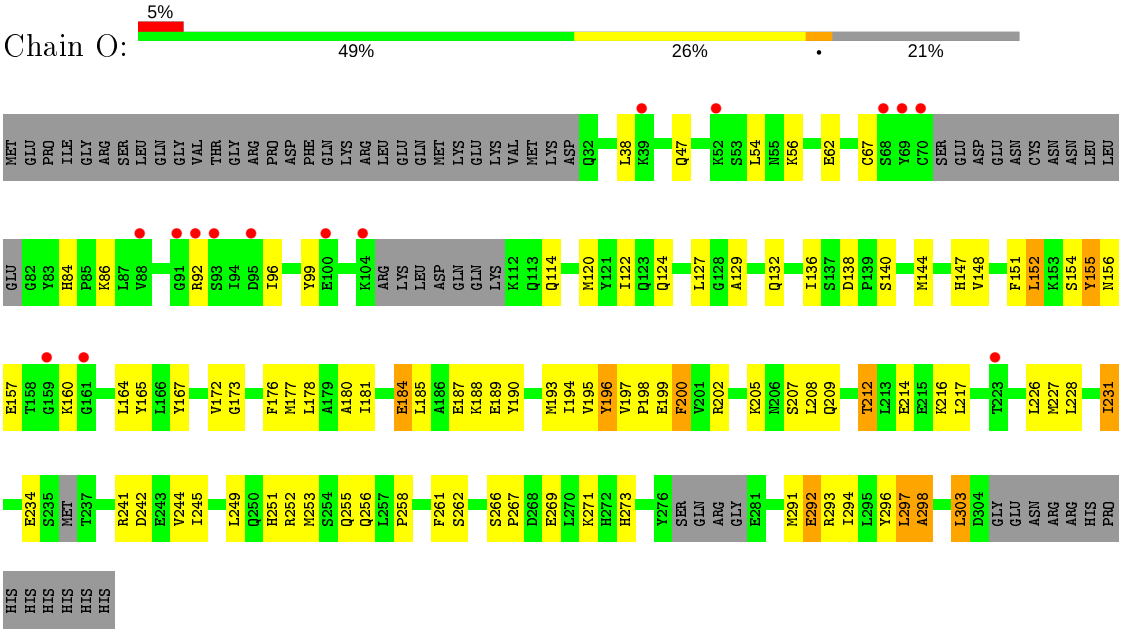
• Molecule 3: Primosomal protein DnaI

Chain N:



• Molecule 3: Primosomal protein DnaI

Chain O:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α, β, γ	229.06Å 229.06Å 364.29Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	20.00 – 6.10 20.00 – 6.10	Depositor EDS
% Data completeness (in resolution range)	72.3 (20.00-6.10) 72.3 (20.00-6.10)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.17	Depositor
$< I/\sigma(I) >$ [1]	0.79 (at 5.92Å)	Xtriage
Refinement program	REFMAC 5.5.0072	Depositor
R, R_{free}	0.379 , 0.392 0.379 , 0.390	Depositor DCC
R_{free} test set	974 reflections (5.15%)	wwPDB-VP
Wilson B-factor (Å^2)	362.7	Xtriage
Anisotropy	0.343	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.44 , 952.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.33$, $< L^2 > = 0.16$	Xtriage
Estimated twinning fraction	0.089 for -h,-k,l	Xtriage
F_o,F_c correlation	0.77	EDS
Total number of atoms	31981	wwPDB-VP
Average B, all atoms (Å^2)	140.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.00% 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	0.40	0/2793	0.67	1/3775 (0.0%)
1	B	0.41	0/2793	0.64	0/3775
1	C	0.42	0/2793	0.64	1/3775 (0.0%)
1	D	0.42	0/2793	0.66	0/3775
1	E	0.41	0/2793	0.64	0/3775
1	F	0.41	0/2793	0.65	0/3775
2	G	0.38	0/1134	0.60	0/1514
2	H	0.40	0/1134	0.64	0/1514
2	I	0.41	0/1134	0.62	0/1514
3	J	0.39	0/2024	0.57	0/2714
3	K	0.38	0/2047	0.57	0/2750
3	L	0.42	0/2043	0.59	0/2740
3	M	0.39	0/2042	0.54	0/2743
3	N	0.40	0/2015	0.55	0/2702
3	O	0.39	0/2066	0.56	0/2776
All	All	0.40	0/32397	0.62	2/43617 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	104	TYR	C-N-CA	5.03	134.27	121.70
1	A	38	LEU	N-CA-CB	5.02	120.44	110.40

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	2763	0	2811	190	0
1	B	2763	0	2811	140	0
1	C	2763	0	2811	197	0
1	D	2763	0	2811	156	0
1	E	2763	0	2811	171	0
1	F	2763	0	2811	202	0
2	G	1122	0	1144	41	0
2	H	1122	0	1144	53	0
2	I	1122	0	1144	55	0
3	J	1992	0	1977	355	0
3	K	2013	0	2003	195	0
3	L	2010	0	1994	456	0
3	M	2008	0	1988	265	0
3	N	1983	0	1973	376	0
3	O	2031	0	2017	71	0
All	All	31981	0	32250	2589	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 40.

The worst 5 of 2589 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:L:33:ASP:HB3	3:L:60:TYR:CE2	1.25	1.63
3:N:45:ILE:HG21	3:N:50:ILE:CG1	1.19	1.62
3:L:86:LYS:CB	3:L:99:TYR:HE1	1.05	1.60
3:N:45:ILE:HD13	3:N:50:ILE:CD1	1.13	1.60
3:L:88:VAL:CG2	3:L:255:GLN:HB2	1.32	1.58

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	348/454 (77%)	243 (70%)	79 (23%)	26 (8%)	1	13
1	B	348/454 (77%)	231 (66%)	95 (27%)	22 (6%)	1	16
1	C	348/454 (77%)	228 (66%)	90 (26%)	30 (9%)	1	11
1	D	348/454 (77%)	234 (67%)	87 (25%)	27 (8%)	1	12
1	E	348/454 (77%)	238 (68%)	82 (24%)	28 (8%)	1	12
1	F	348/454 (77%)	222 (64%)	96 (28%)	30 (9%)	1	11
2	G	136/143 (95%)	100 (74%)	28 (21%)	8 (6%)	1	17
2	H	136/143 (95%)	88 (65%)	40 (29%)	8 (6%)	1	17
2	I	136/143 (95%)	99 (73%)	32 (24%)	5 (4%)	3	24
3	J	233/317 (74%)	166 (71%)	55 (24%)	12 (5%)	2	19
3	K	238/317 (75%)	175 (74%)	47 (20%)	16 (7%)	1	15
3	L	235/317 (74%)	162 (69%)	60 (26%)	13 (6%)	2	18
3	M	237/317 (75%)	172 (73%)	52 (22%)	13 (6%)	2	18
3	N	232/317 (73%)	179 (77%)	41 (18%)	12 (5%)	2	19
3	O	240/317 (76%)	176 (73%)	57 (24%)	7 (3%)	4	29
All	All	3911/5055 (77%)	2713 (69%)	941 (24%)	257 (7%)	1	15

5 of 257 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	97	PRO
1	A	203	SER
1	A	259	ASN
1	A	294	PRO
1	A	325	ILE

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	298/386 (77%)	260 (87%)	38 (13%)	4	18
1	B	298/386 (77%)	264 (89%)	34 (11%)	5	21
1	C	298/386 (77%)	263 (88%)	35 (12%)	5	21
1	D	298/386 (77%)	260 (87%)	38 (13%)	4	18
1	E	298/386 (77%)	260 (87%)	38 (13%)	4	18
1	F	298/386 (77%)	263 (88%)	35 (12%)	5	21
2	G	117/116 (101%)	105 (90%)	12 (10%)	7	25
2	H	117/116 (101%)	109 (93%)	8 (7%)	16	41
2	I	117/116 (101%)	102 (87%)	15 (13%)	4	18
3	J	221/289 (76%)	199 (90%)	22 (10%)	7	26
3	K	225/289 (78%)	197 (88%)	28 (12%)	4	19
3	L	223/289 (77%)	203 (91%)	20 (9%)	9	30
3	M	224/289 (78%)	208 (93%)	16 (7%)	14	39
3	N	220/289 (76%)	199 (90%)	21 (10%)	8	28
3	O	227/289 (78%)	210 (92%)	17 (8%)	13	38
All	All	3479/4398 (79%)	3102 (89%)	377 (11%)	6	23

5 of 377 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	267	LYS
1	F	355	VAL
3	N	200	PHE
1	E	324	LEU
1	F	130	TYR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 62 such sidechains are listed below:

Mol	Chain	Res	Type
2	G	462	ASN
3	J	84	HIS
3	N	133	GLN
2	H	494	HIS
3	J	183	ASN

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 carbohydrates 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	358/454 (78%)	-0.48	2 (0%) 89 84	81, 134, 167, 169	0
1	B	358/454 (78%)	-0.52	1 (0%) 94 90	100, 138, 156, 171	0
1	C	358/454 (78%)	-0.50	1 (0%) 94 90	87, 123, 164, 174	0
1	D	358/454 (78%)	-0.48	0 100 100	81, 130, 191, 194	0
1	E	358/454 (78%)	-0.51	2 (0%) 89 84	65, 124, 160, 176	0
1	F	358/454 (78%)	-0.41	0 100 100	77, 142, 174, 175	0
2	G	132/143 (92%)	-0.63	0 100 100	106, 130, 143, 145	0
2	H	132/143 (92%)	-0.74	0 100 100	78, 120, 134, 135	0
2	I	132/143 (92%)	-0.58	0 100 100	107, 128, 149, 149	0
3	J	245/317 (77%)	0.19	7 (2%) 51 44	120, 170, 188, 195	0
3	K	248/317 (78%)	-0.06	3 (1%) 79 71	120, 171, 186, 200	0
3	L	247/317 (77%)	0.07	5 (2%) 65 58	120, 172, 201, 204	0
3	M	247/317 (77%)	0.05	9 (3%) 42 38	120, 163, 178, 182	0
3	N	244/317 (76%)	0.17	9 (3%) 41 37	120, 180, 194, 196	0
3	O	250/317 (78%)	0.04	15 (6%) 21 21	120, 160, 178, 179	0
All	All	4025/5055 (79%)	-0.29	54 (1%) 77 68	65, 140, 186, 204	0

The worst 5 of 54 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	203	SER	4.7
3	J	63	GLN	4.3
3	N	95	ASP	3.6
3	J	64	SER	3.5
1	E	260	ALA	3.3

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

There are no carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.