wwPDB X-ray Structure Validation Summary Report (i)

Sep 24, 2023 – 08:51 AM EDT

PDB ID	:	5UJ8
Title	:	Human Origin Recognition Complex subunits 2 and 3
Authors	:	Tocilj, A.; On, K.F.; Elkayam, E.; Joshua-Tor, L.
Deposited on	:	2017-01-17
Resolution	:	6.00 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:

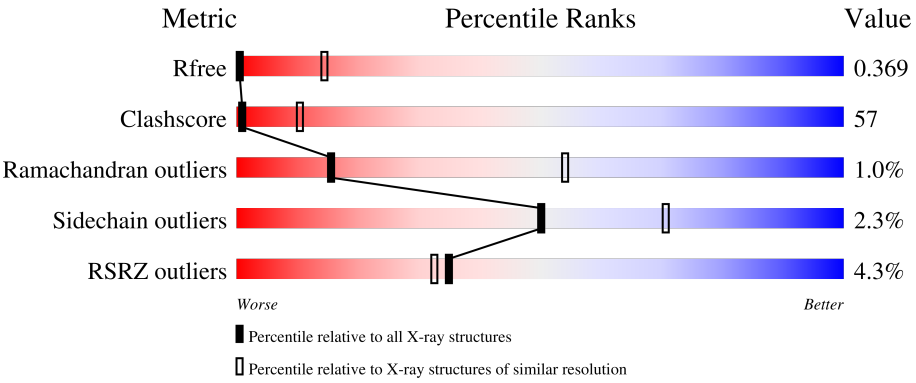
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 6.00 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	1000 (8.00-3.88)
Clashscore	141614	1049 (8.00-3.90)
Ramachandran outliers	138981	1016 (8.00-3.86)
Sidechain outliers	138945	1017 (8.00-3.82)
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 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	712	3% 24% 50% . 22%
1	B	712	3% 25% 50% . 22%
1	C	712	5% 23% 52% . 22%
1	D	712	4% 26% 48% . 22%
2	E	347	% 22% 29% . 47%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	347	2% 22% 29% 47%
2	G	347	3% 22% 29% 47%
2	H	347	% 21% 31% 47%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 24144 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 Origin recognition complex subunit 3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	553	Total	C	N	O	S	0	0	0
			4524	2920	767	812	25			
1	B	553	Total	C	N	O	S	0	0	0
			4524	2920	767	812	25			
1	C	553	Total	C	N	O	S	0	0	0
			4524	2920	767	812	25			
1	D	553	Total	C	N	O	S	0	0	0
			4524	2920	767	812	25			

- Molecule 2 is a protein called Origin recognition complex subunit 2.

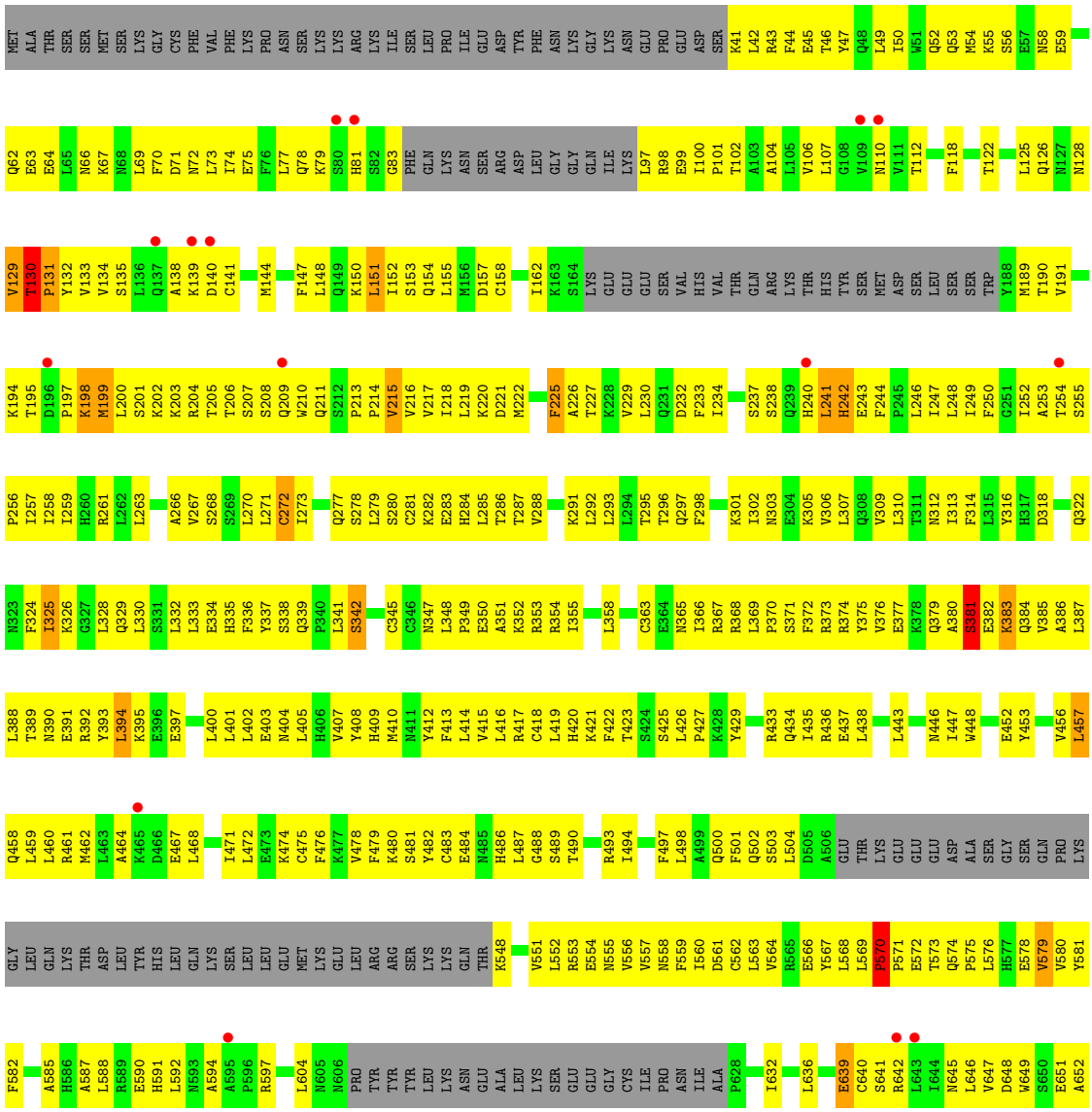
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	183	Total	C	N	O	S	0	0	0
			1512	978	249	280	5			
2	F	183	Total	C	N	O	S	0	0	0
			1512	978	249	280	5			
2	G	183	Total	C	N	O	S	0	0	0
			1512	978	249	280	5			
2	H	183	Total	C	N	O	S	0	0	0
			1512	978	249	280	5			

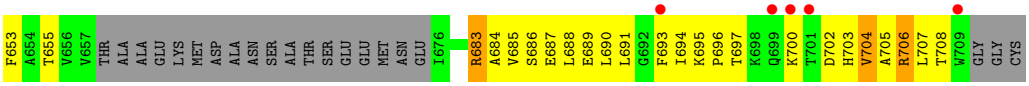
3 Residue-property plots (i)

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: Origin recognition complex subunit 3

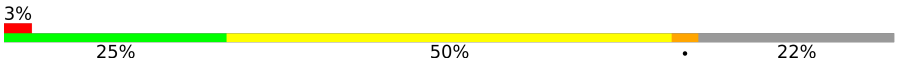
Chain A:

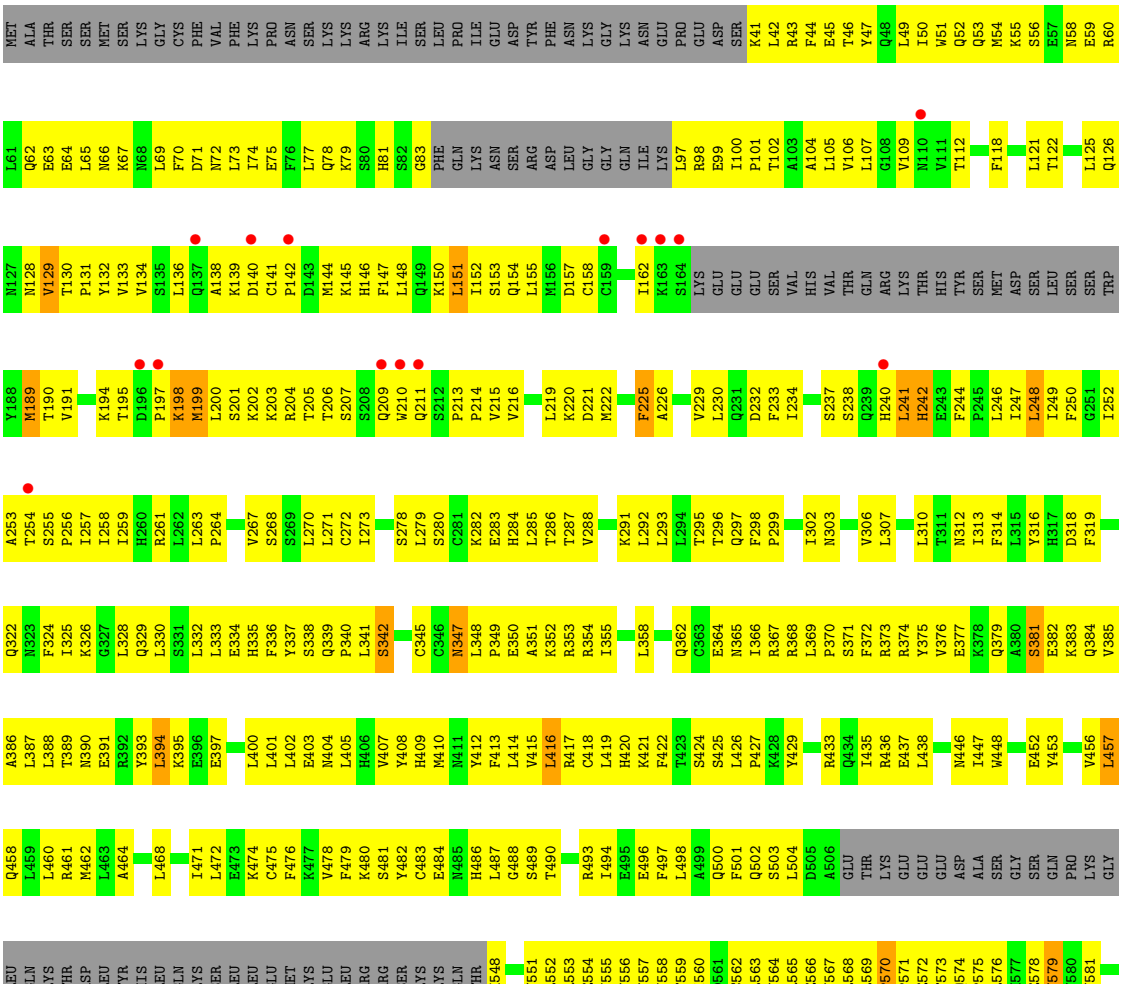




• Molecule 1: Origin recognition complex subunit 3

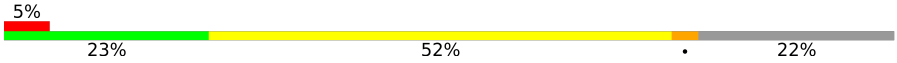
Chain B:

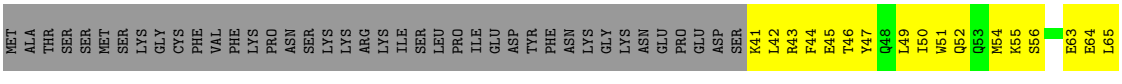


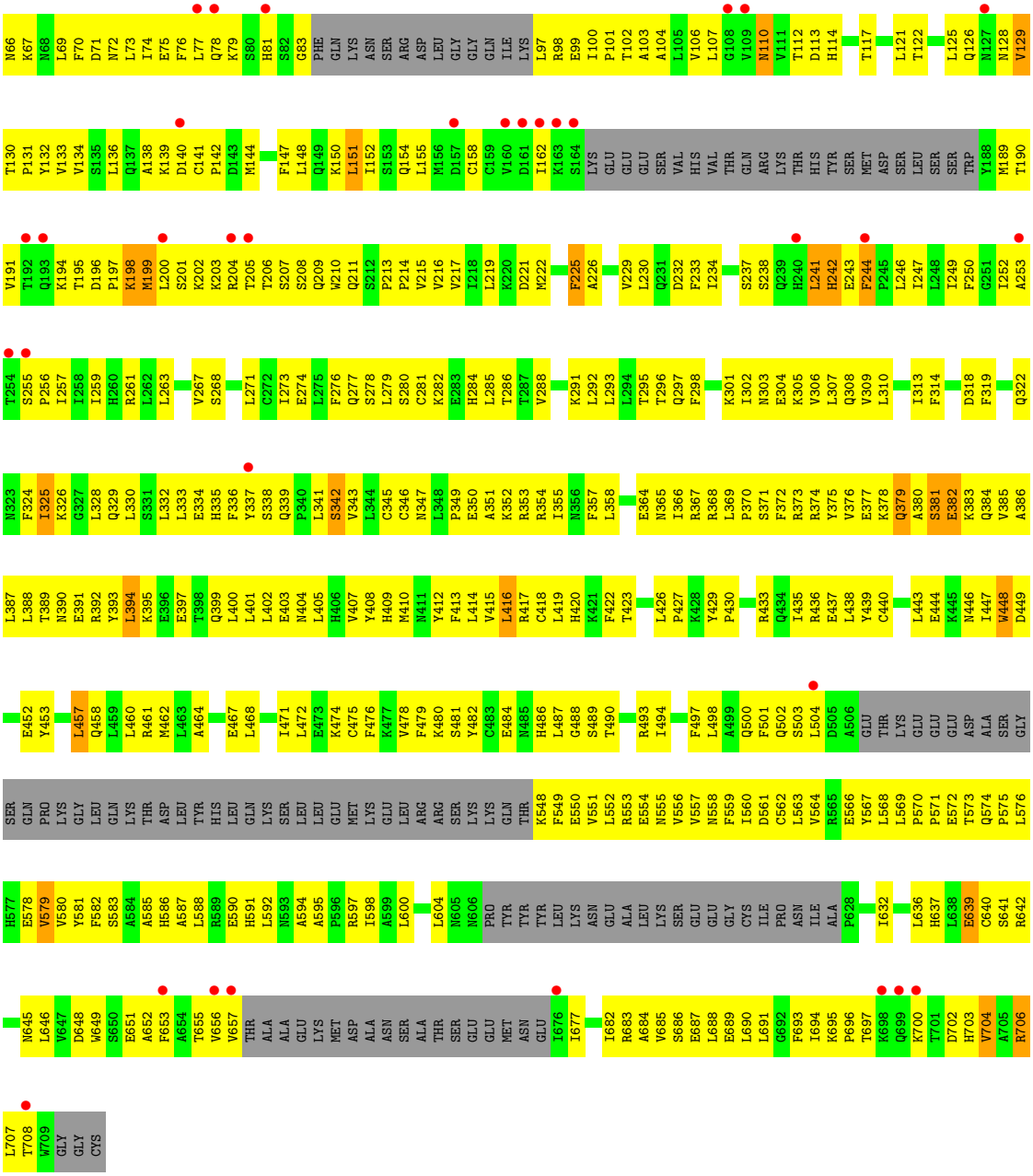


• Molecule 1: Origin recognition complex subunit 3

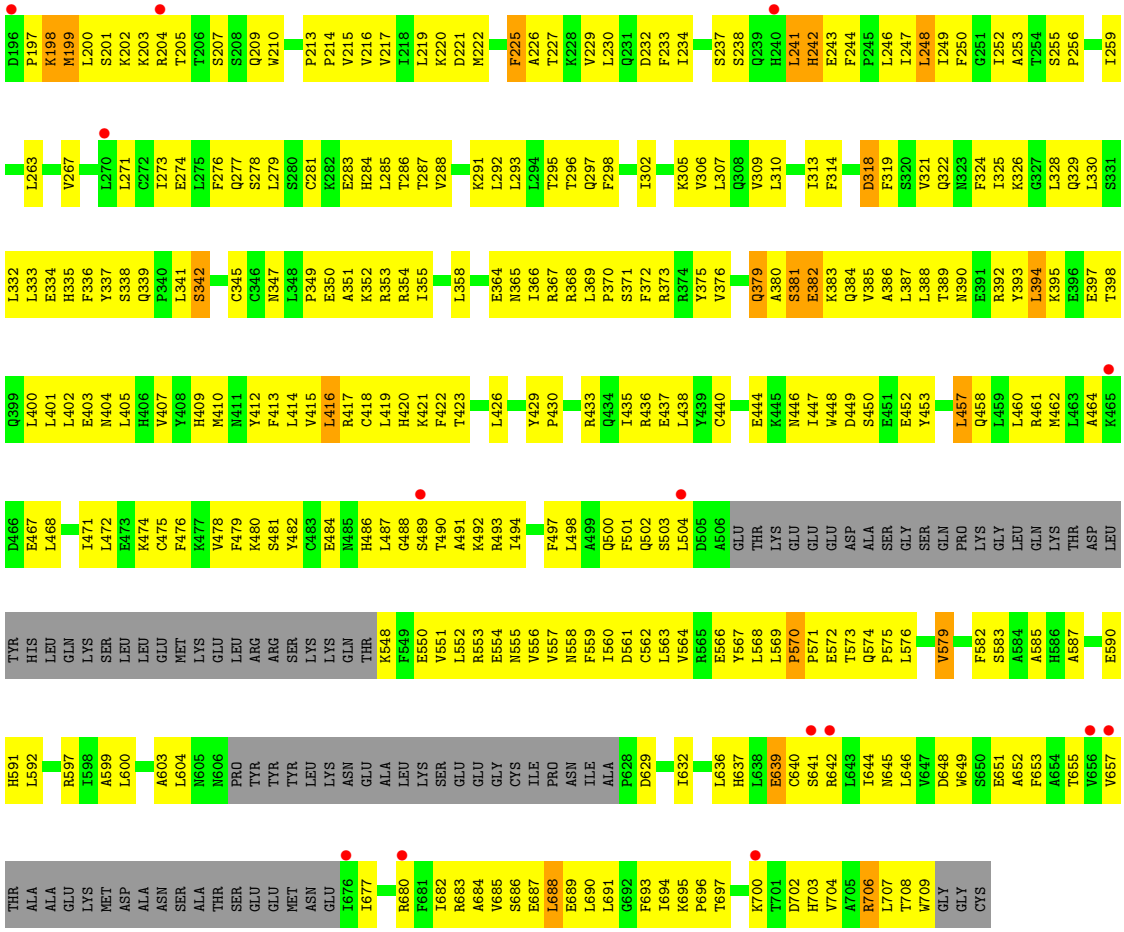
Chain C:





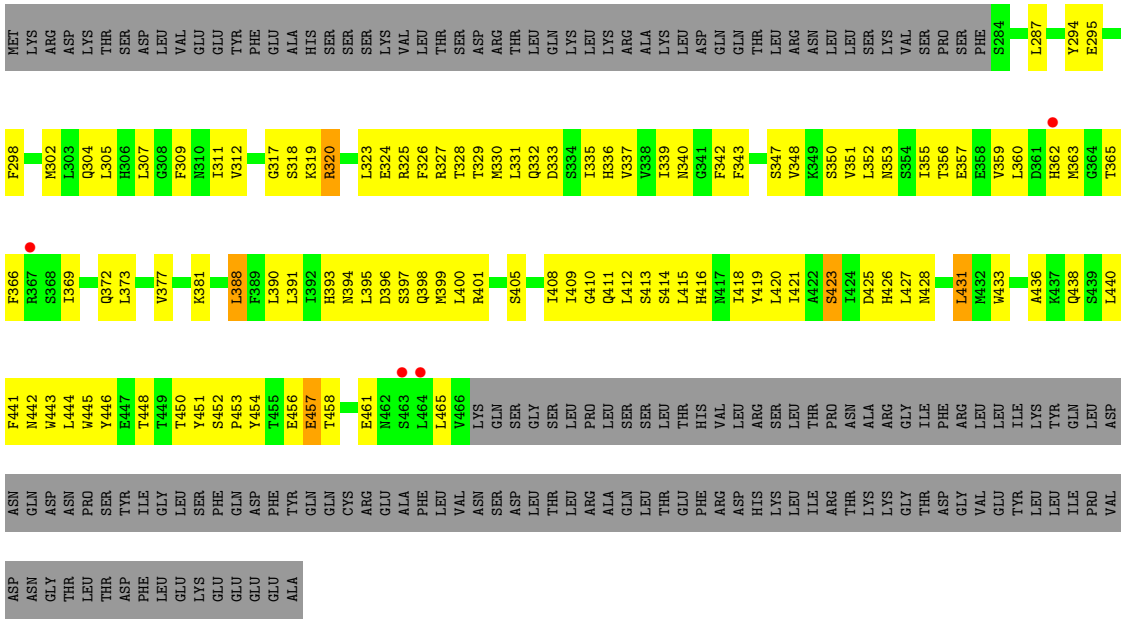


• Molecule 1: Origin recognition complex subunit 3



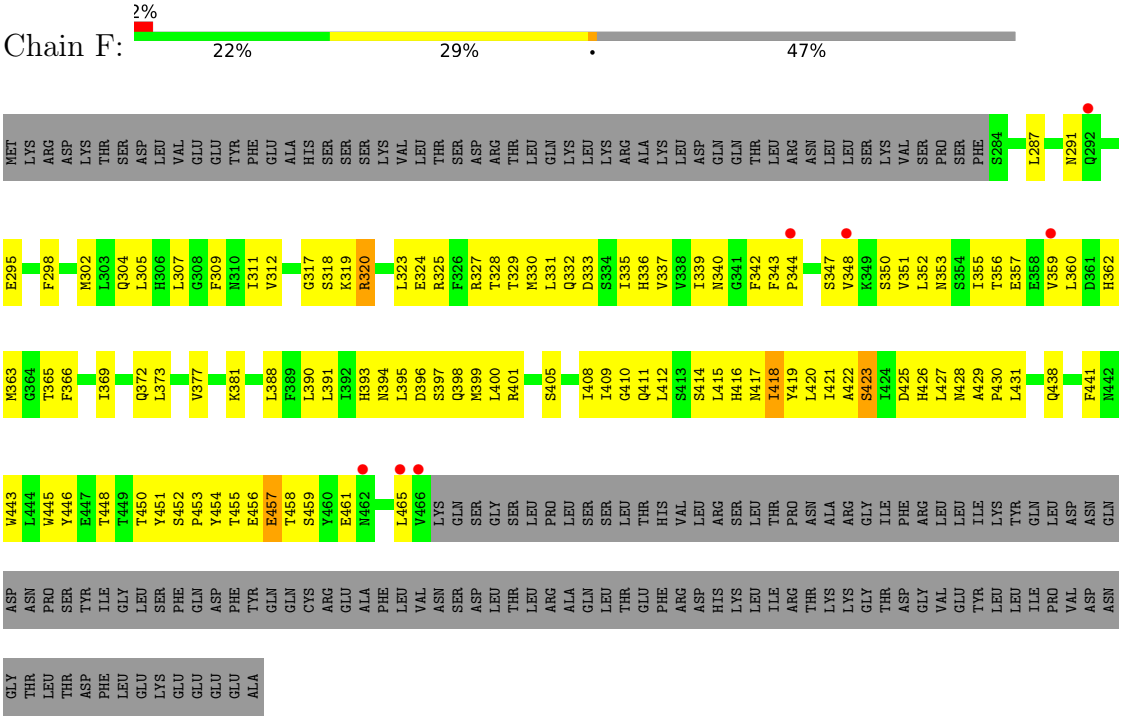
• Molecule 2: Origin recognition complex subunit 2

Chain E:



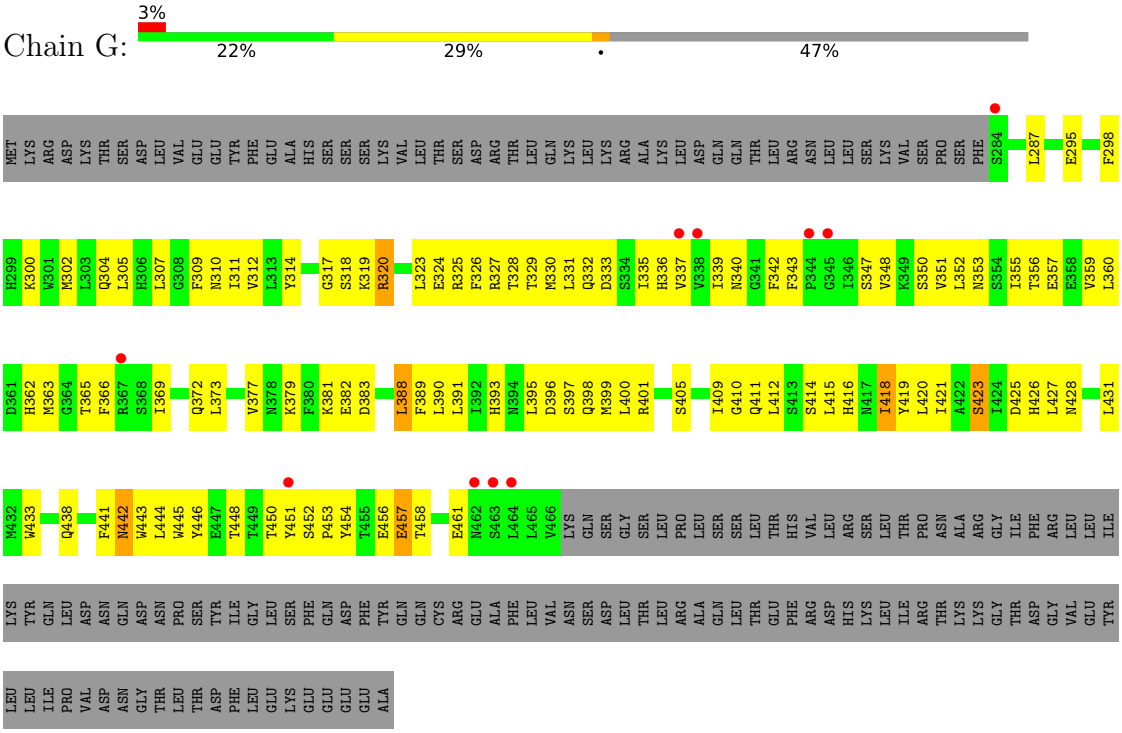
• Molecule 2: Origin recognition complex subunit 2

Chain F:



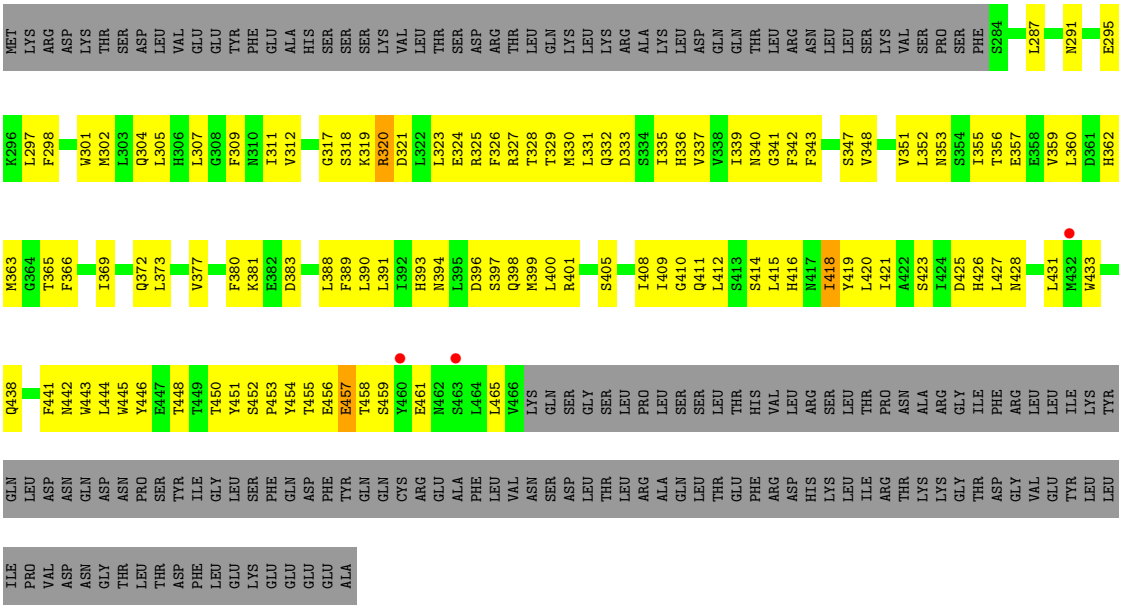
• Molecule 2: Origin recognition complex subunit 2

Chain G:



• Molecule 2: Origin recognition complex subunit 2

Chain H:


MET LYS ARG ASP LYS THR SER ASP LEU VAL GLU GLU TYR PHE GLU ALA HIS SER SER SER LYS VAL LEU THR SER ASP ARG THR LEU GLN LYS LEU LYS ARG ALA LYS LEU ASP GLN GLN THR LEU ARG ASN LEU LEU SER LYS VAL SER PRO SER PHE S284 L287 N291 E295
K296 L297 F298 W301 M302 L303 Q304 L305 H306 L307 G308 F309 N310 I311 V312 G317 S318 K319 R320 D321 L322 L323 E324 R325 F326 R327 T328 T329 M330 L331 Q332 D333 S334 I335 H336 V337 V338 I339 N340 G341 F342 F343 S347 V348 V351 L352 N353 S354 I355 T356 E357 E358 V359 L360 D361 H362
M363 G364 T365 F366 I369 Q372 L373 V377 F380 K381 E382 D383 L388 F389 L390 L391 I392 H393 N394 L395 D396 S397 Q398 M399 L400 R401 S405 I408 I409 G410 Q411 L412 S413 S414 L415 H416 N417 I418 Y419 L420 I421 A422 S423 I424 D425 H426 L427 N428 L431 M432 W433
Q438 F441 N442 W443 L444 W445 Y446 E447 T448 T449 T450 Y451 S452 P453 Y454 T455 E456 E457 T458 S459 Y460 E461 N462 S463 L464 L465 V466 LYS GLN SER GLY SER LEU PRO LEU SER SER LEU THR HIS VAL LEU ARG SER LEU THR PRO ASN ALA ARG GLY ILE PHE ARG LEU LEU ILE LYS TYR
GLN LEU ASP ASN GLN ASP ASN PRO SER TYR ILE GLY LEU SER PHE GLN ASP PHE TYR GLN GLN CYS ARG GLU ALA PHE LEU VAL ASN SER ASP LEU THR LEU ARG ALA GLN LEU THR GLU PHE ARG ASP HIS LYS LEU ILE ARG THR LYS LYS GLY THR ASP GLY VAL GLU TYR LEU LEU
ILE PRO VAL ASP ASN GLY THR LEU THR ASP PHE LEU GLU LYS GLU GLU GLU GLU ALA

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	87.26Å 114.96Å 316.45Å 90.00° 90.72° 90.00°	Depositor
Resolution (Å)	20.07 – 6.00 20.07 – 6.00	Depositor EDS
% Data completeness (in resolution range)	97.8 (20.07-6.00) 94.2 (20.07-6.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.50 (at 5.93Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, R_{free}	0.318 , 0.368 0.336 , 0.369	Depositor DCC
R_{free} test set	753 reflections (4.96%)	wwPDB-VP
Wilson B-factor (Å^2)	287.3	Xtriage
Anisotropy	0.393	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.25 , 190.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.42$, $< L^2 > = 0.24$	Xtriage
Estimated twinning fraction	0.155 for h,-k,-l	Xtriage
F_o,F_c correlation	0.86	EDS
Total number of atoms	24144	wwPDB-VP
Average B, all atoms (Å^2)	303.0	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 28.14 % 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 1.9421e-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

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	0.45	0/4616	0.70	5/6244 (0.1%)
1	B	0.45	0/4616	0.69	4/6244 (0.1%)
1	C	1.12	9/4616 (0.2%)	0.75	7/6244 (0.1%)
1	D	0.48	0/4616	0.71	5/6244 (0.1%)
2	E	0.42	0/1548	0.70	2/2097 (0.1%)
2	F	0.39	0/1548	0.68	1/2097 (0.0%)
2	G	0.40	0/1548	0.69	2/2097 (0.1%)
2	H	0.41	0/1548	0.69	1/2097 (0.0%)
All	All	0.63	9/24656 (0.0%)	0.71	27/33364 (0.1%)

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	5
1	B	0	5
1	C	0	3
1	D	0	3
All	All	0	16

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	199	MET	CG-SD	38.31	2.80	1.81
1	C	244	PHE	CE1-CZ	28.73	1.92	1.37
1	C	244	PHE	CE2-CZ	27.23	1.89	1.37
1	C	244	PHE	CD2-CE2	25.67	1.90	1.39
1	C	244	PHE	CD1-CE1	23.71	1.86	1.39

The worst 5 of 27 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	199	MET	CG-SD-CE	16.48	126.57	100.20
1	C	151	LEU	CA-CB-CG	10.74	140.00	115.30
1	D	151	LEU	CA-CB-CG	10.43	139.29	115.30
1	B	151	LEU	CA-CB-CG	10.19	138.75	115.30
2	H	320	ARG	NE-CZ-NH2	-10.18	115.21	120.30

There are no chirality outliers.

5 of 16 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	130	THR	Peptide
1	A	140	ASP	Peptide
1	A	240	HIS	Peptide
1	A	242	HIS	Peptide
1	A	639	GLU	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	4524	0	4619	564	0
1	B	4524	0	4619	544	0
1	C	4524	0	4619	578	0
1	D	4524	0	4619	547	0
2	E	1512	0	1495	153	0
2	F	1512	0	1495	150	0
2	G	1512	0	1495	157	0
2	H	1512	0	1495	163	0
All	All	24144	0	24456	2749	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 57.

The worst 5 of 2749 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:C:244:PHE:CE1	1:C:244:PHE:CD1	1.86	1.61

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:244:PHE:CD2	1:C:244:PHE:CE2	1.90	1.59
1:C:244:PHE:CE1	1:C:244:PHE:CZ	1.91	1.57
1:C:244:PHE:CD2	1:C:246:LEU:HG	1.41	1.48
1:C:202:LYS:NZ	1:C:244:PHE:CE1	1.73	1.40

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	541/712 (76%)	460 (85%)	72 (13%)	9 (2%)	9 42
1	B	541/712 (76%)	463 (86%)	72 (13%)	6 (1%)	14 52
1	C	541/712 (76%)	462 (85%)	72 (13%)	7 (1%)	12 48
1	D	541/712 (76%)	464 (86%)	70 (13%)	7 (1%)	12 48
2	E	181/347 (52%)	161 (89%)	20 (11%)	0	100 100
2	F	181/347 (52%)	161 (89%)	20 (11%)	0	100 100
2	G	181/347 (52%)	161 (89%)	20 (11%)	0	100 100
2	H	181/347 (52%)	163 (90%)	18 (10%)	0	100 100
All	All	2888/4236 (68%)	2495 (86%)	364 (13%)	29 (1%)	15 54

5 of 29 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	129	VAL
1	A	130	THR
1	A	383	LYS
1	B	129	VAL
1	B	383	LYS

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	517/659 (78%)	504 (98%)	13 (2%)	47	68
1	B	517/659 (78%)	505 (98%)	12 (2%)	50	71
1	C	517/659 (78%)	506 (98%)	11 (2%)	53	72
1	D	517/659 (78%)	505 (98%)	12 (2%)	50	71
2	E	171/323 (53%)	168 (98%)	3 (2%)	59	77
2	F	171/323 (53%)	168 (98%)	3 (2%)	59	77
2	G	171/323 (53%)	166 (97%)	5 (3%)	42	64
2	H	171/323 (53%)	168 (98%)	3 (2%)	59	77
All	All	2752/3928 (70%)	2690 (98%)	62 (2%)	50	71

5 of 62 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	342	SER
2	G	418	ILE
1	C	706	ARG
2	G	388	LEU
2	H	418	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	384	GLN
2	F	435	HIS
2	H	442	ASN
1	B	209	GLN
1	B	242	HIS

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)

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	553/712 (77%)	0.03	20 (3%) 42 38	214, 295, 363, 401	0
1	B	553/712 (77%)	0.12	21 (3%) 40 36	208, 300, 363, 410	0
1	C	553/712 (77%)	0.27	33 (5%) 21 21	219, 302, 365, 407	0
1	D	553/712 (77%)	0.22	29 (5%) 27 27	227, 299, 363, 406	0
2	E	183/347 (52%)	-0.02	4 (2%) 62 54	256, 307, 348, 381	0
2	F	183/347 (52%)	0.11	7 (3%) 40 36	254, 310, 353, 374	0
2	G	183/347 (52%)	0.07	10 (5%) 25 25	250, 306, 352, 375	0
2	H	183/347 (52%)	0.06	3 (1%) 72 64	248, 310, 349, 376	0
All	All	2944/4236 (69%)	0.13	127 (4%) 35 32	208, 303, 360, 410	0

The worst 5 of 127 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	164	SER	7.9
1	C	164	SER	5.1
1	D	163	LYS	5.0
2	G	463	SER	4.9
1	C	81	HIS	4.4

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

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.