wwPDB EM Map/Model Validation Report (i)

Apr 10, 2016 – 01:59 PM BST

PDB ID	:	3JC5
EMDB ID:	:	EMD-6535
Title	:	Structure of the eukaryotic replicative CMG helicase and pumpjack motion
Authors	:	Li, H.; Bai, L.; Yuan, Z.; Sun, J.; Georgescu, R.E.; Liu, J.; O' Donnell, M.E.
Deposited on	:	2015-11-24
Resolution	:	4.70 Å(reported)
Based on PDB ID	:	2Q9Q

This is a wwPDB EM Map/Model Validation Report for a publicly released PDB/EMDB entry.
For rigid body fitted models, validation errors reported here could
stem from errors in the original structure(s) used in the fitting.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
http://wwpdb.org/validation/2016/EMValidationReportHelp

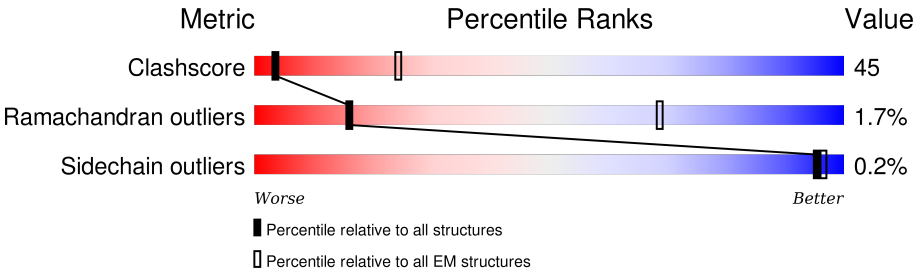
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk27241

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is 4.70 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.



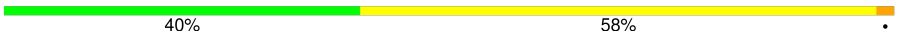
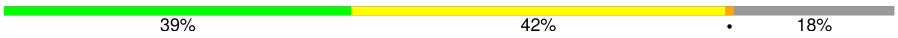
Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	114402	924
Ramachandran outliers	111179	726
Sidechain outliers	111093	686

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. 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%

Mol	Chain	Length	Quality of chain
1	2	868	34% 32% • 33%
2	3	971	27% 32% • 39%
3	4	933	28% 31% • 39%
4	5	775	37% 44% • 16%
5	6	1017	31% 33% • 34%
6	7	845	36% 40% • 23%
7	c	650	84% • 15%
8	D	294	38% 35% • 25%
9	B	213	41% 43% 15%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
10	A	208	
11	C	194	

2 Entry composition

There are 11 unique types of molecules in this entry. The entry contains 40041 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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 replication licensing factor MCM2.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	2	584	4600	2904	819	861	16	0	0

- Molecule 2 is a protein called DNA replication licensing factor MCM3.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	3	588	4613	2909	820	871	13	0	0

- Molecule 3 is a protein called DNA replication licensing factor MCM4.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
3	4	569	4516	2842	783	864	27	0	0

- Molecule 4 is a protein called Minichromosome maintenance protein 5.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
4	5	653	5171	3251	896	1001	23	0	0

- Molecule 5 is a protein called DNA replication licensing factor MCM6.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
5	6	671	5211	3291	916	981	23	0	0

- Molecule 6 is a protein called DNA replication licensing factor MCM7.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
6	7	652	5148	3249	895	977	27	0	0

- Molecule 7 is a protein called Cell division control protein 45.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	c	553	Total	C	N	O	S	0	0
			4470	2852	759	846	13		

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
c	22	ALA	HIS	CONFLICT	UNP Q08032
c	155	GLU	GLN	CONFLICT	UNP Q08032
c	551	THR	TRP	CONFLICT	UNP Q08032

- Molecule 8 is a protein called DNA replication complex GINS protein SLD5.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	D	221	Total	C	N	O	S	0	0
			1820	1159	300	348	13		

- Molecule 9 is a protein called DNA replication complex GINS protein PSF2.

Mol	Chain	Residues	Atoms					AltConf	Trace
9	B	181	Total	C	N	O	S	0	0
			1513	978	261	270	4		

- Molecule 10 is a protein called DNA replication complex GINS protein PSF1.

Mol	Chain	Residues	Atoms					AltConf	Trace
10	A	208	Total	C	N	O	S	0	0
			1691	1062	287	332	10		

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	161	ALA	VAL	CONFLICT	UNP Q12488
A	192	GLN	ARG	CONFLICT	UNP Q12488
A	207	LEU	LYS	CONFLICT	UNP Q12488

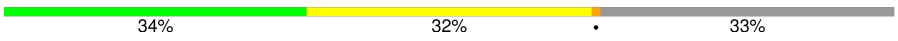
- Molecule 11 is a protein called DNA replication complex GINS protein PSF3.

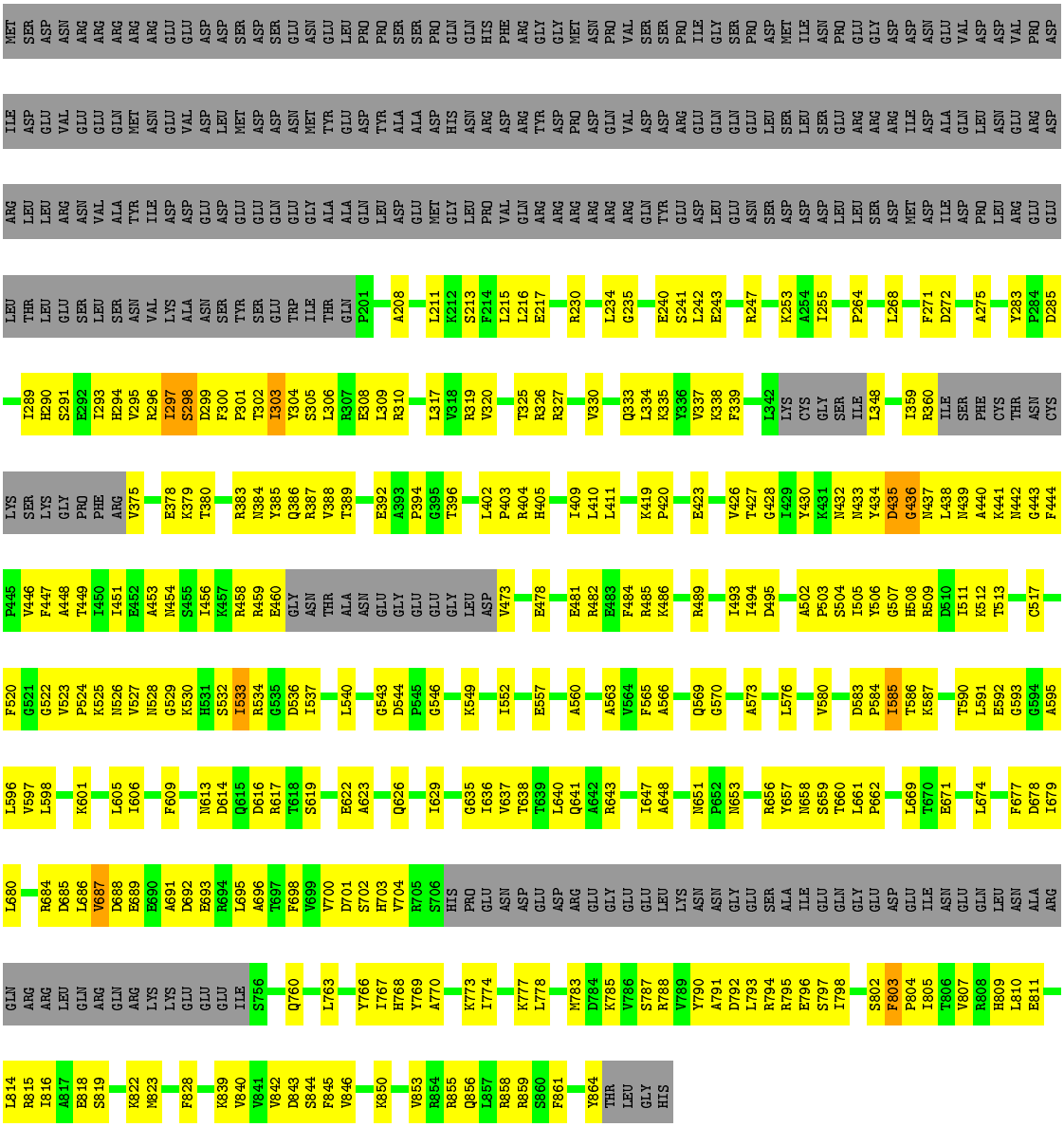
Mol	Chain	Residues	Atoms					AltConf	Trace
11	C	159	Total	C	N	O	S	0	0
			1288	843	207	232	6		

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 errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. 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. 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.

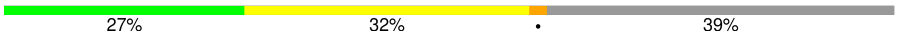
- Molecule 1: DNA replication licensing factor MCM2

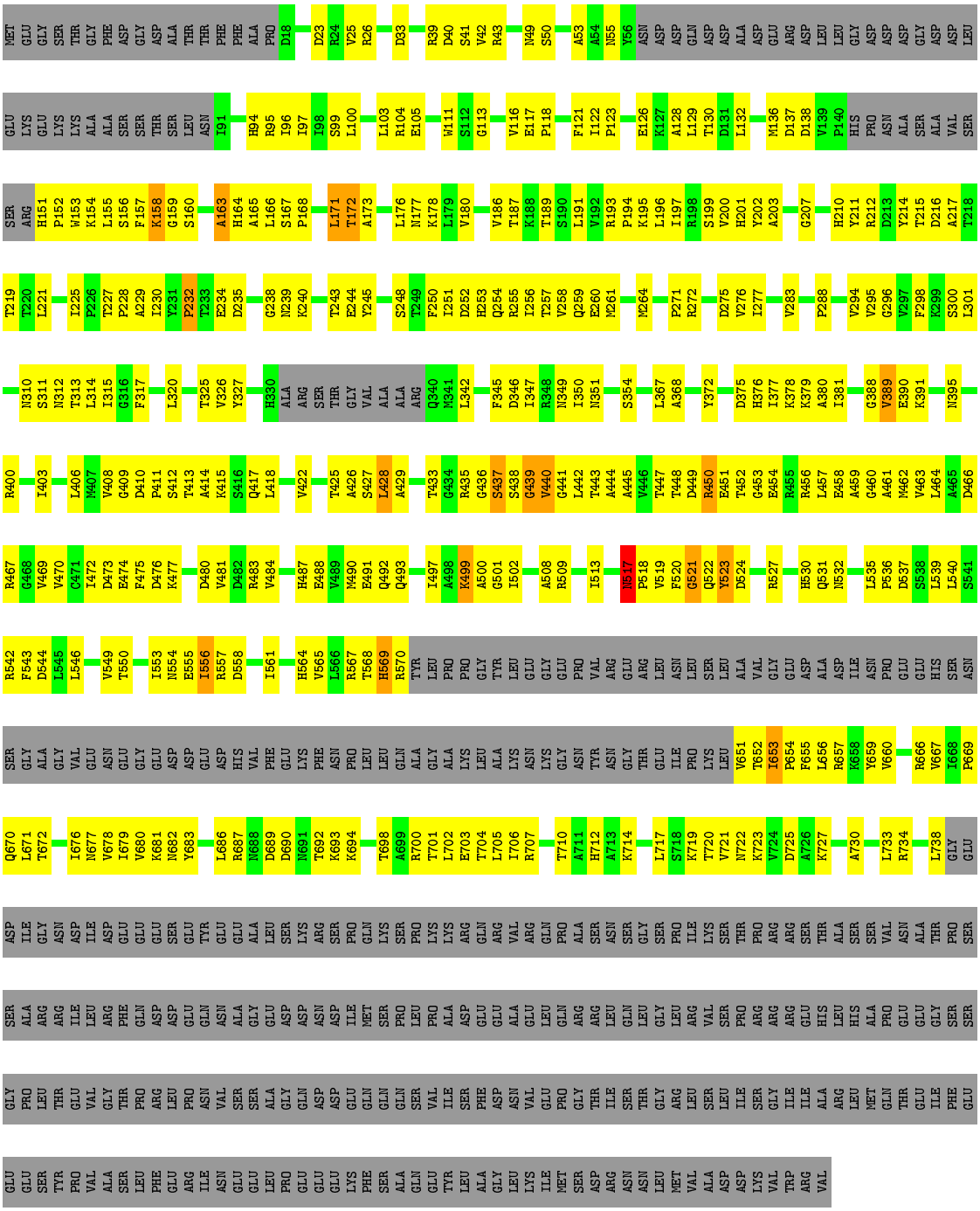
Chain 2: 



- Molecule 2: DNA replication licensing factor MCM3

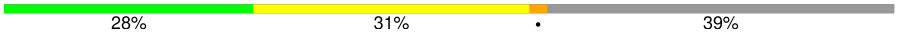
Chain 3:

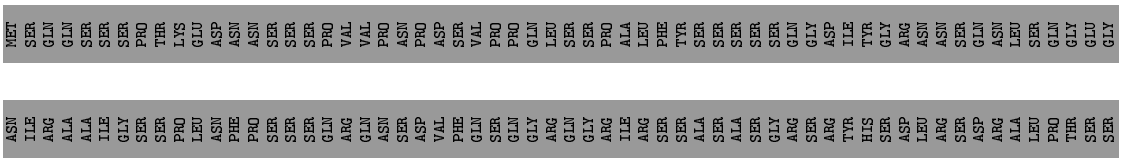


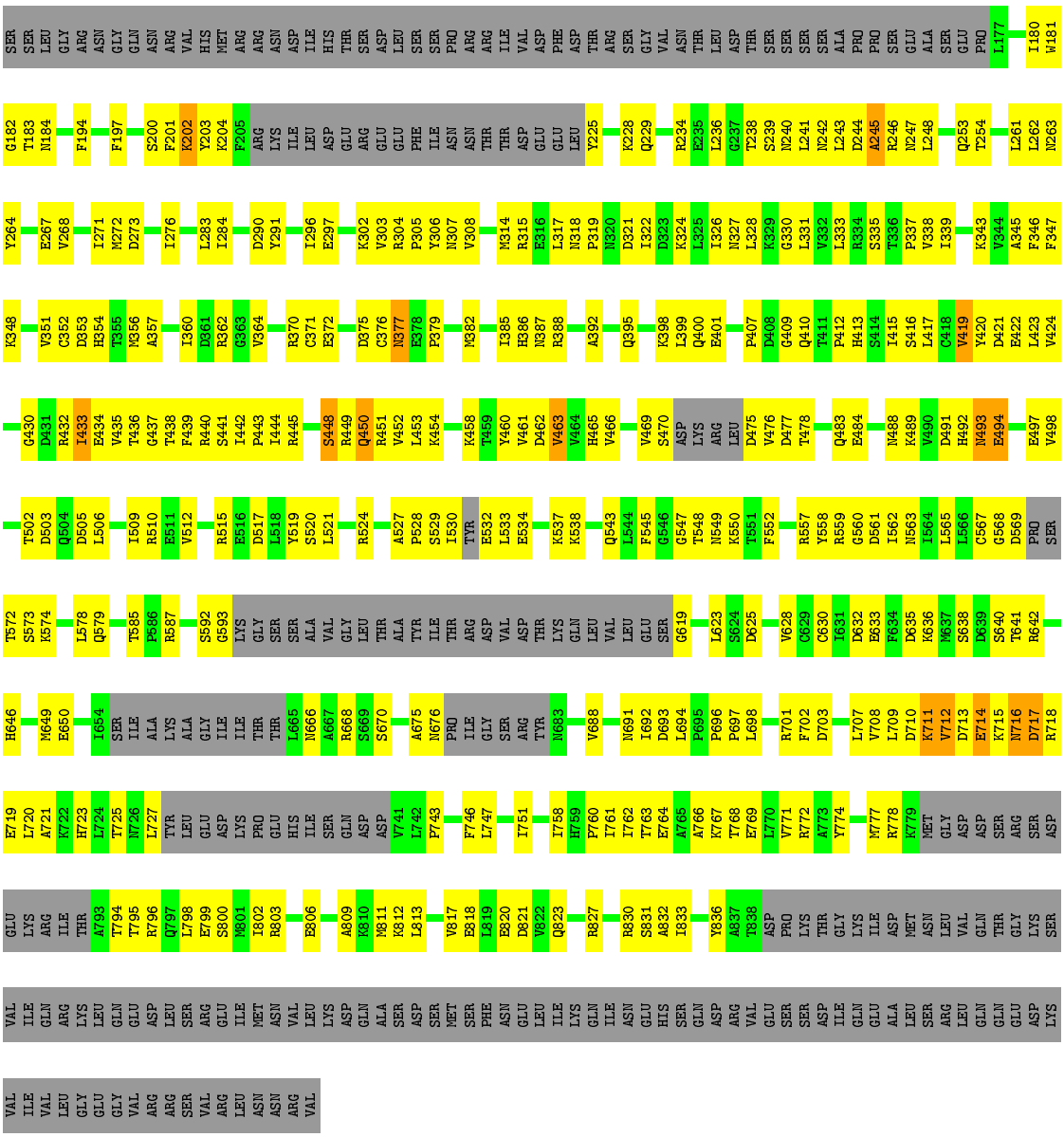


• Molecule 3: DNA replication licensing factor MCM4

Chain 4:

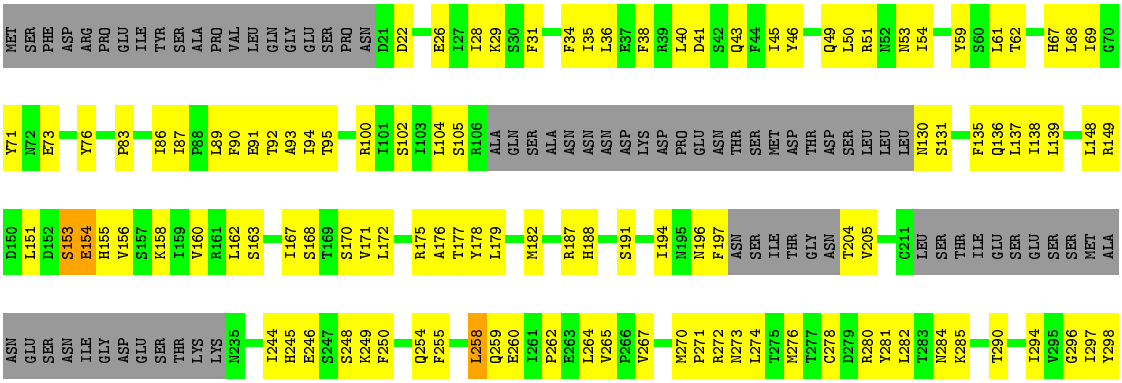


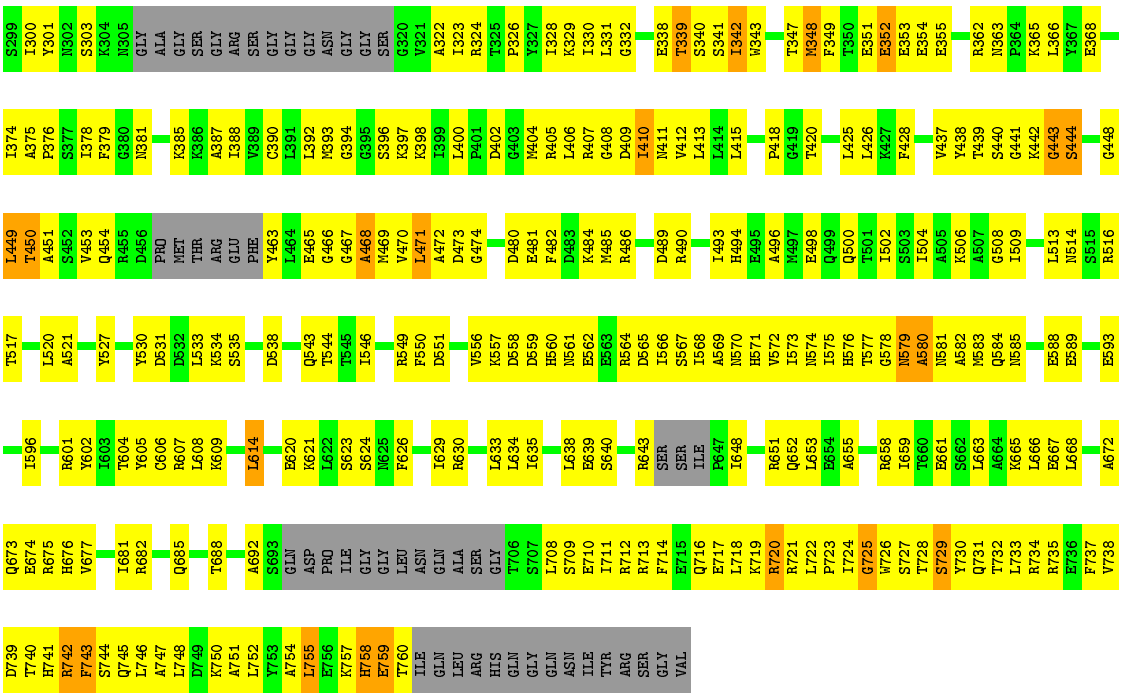




- Molecule 4: Minichromosome maintenance protein 5

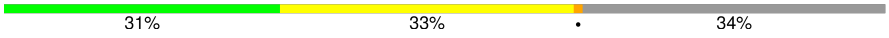
Chain 5:

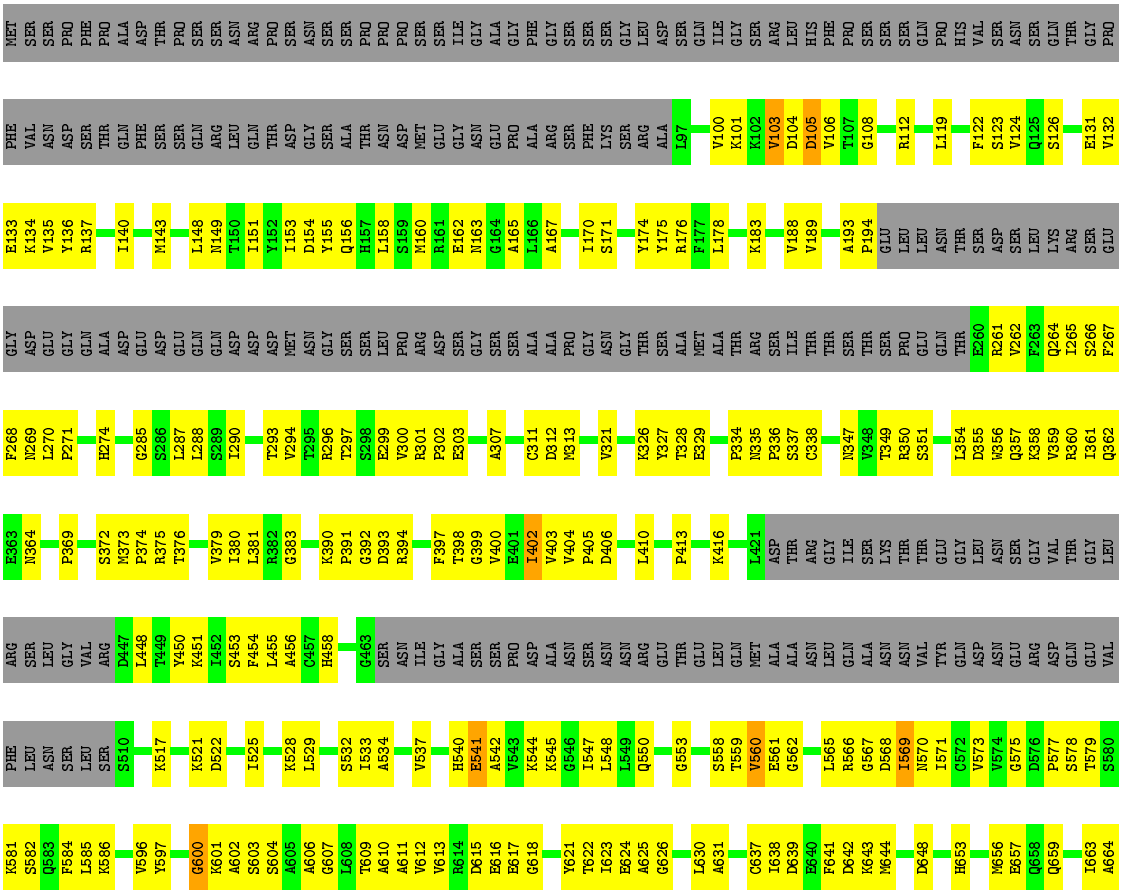


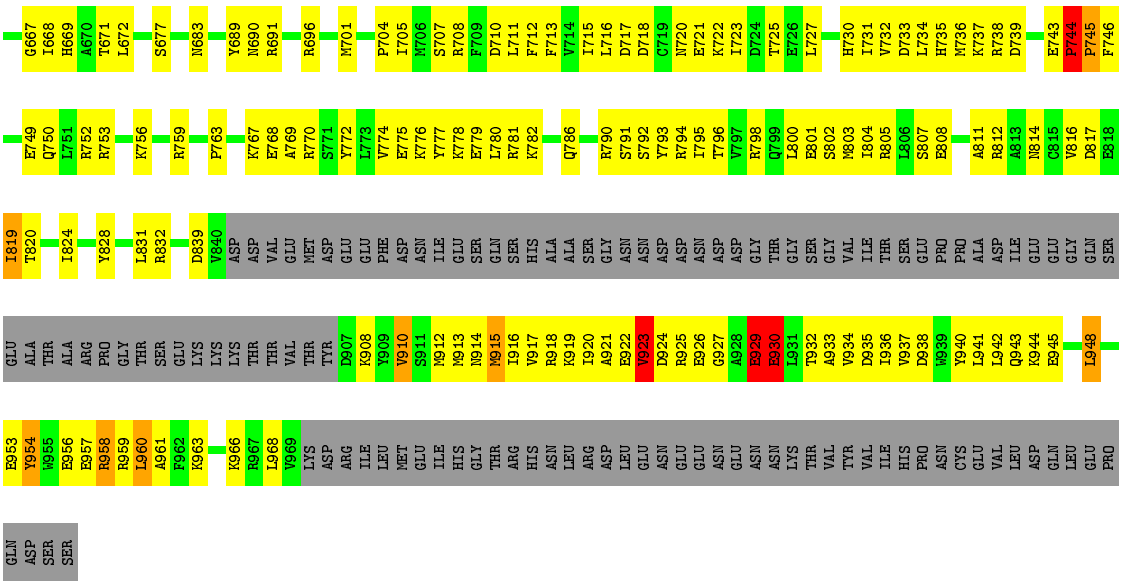


• Molecule 5: DNA replication licensing factor MCM6

Chain 6:

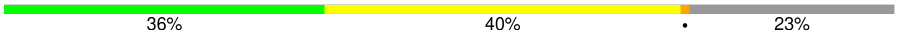


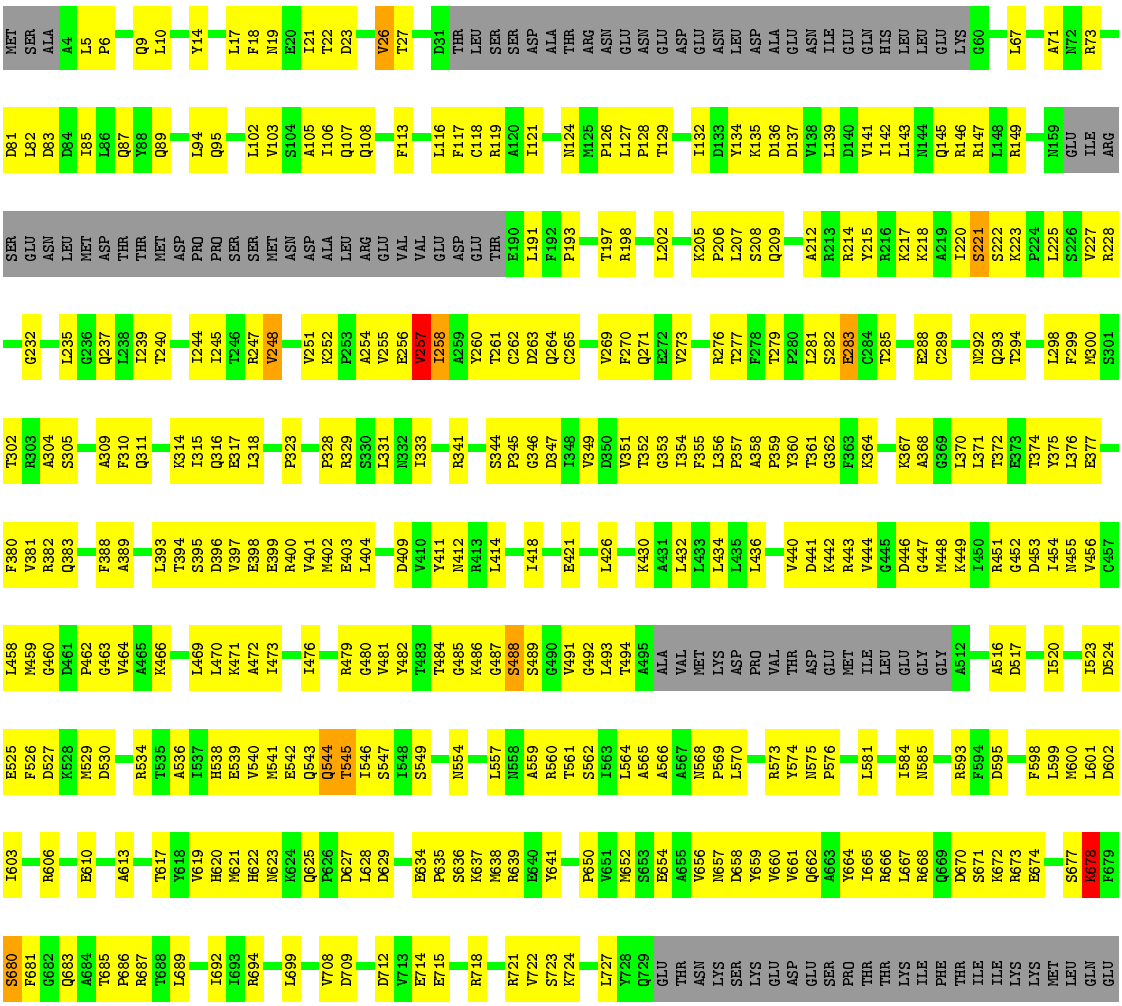


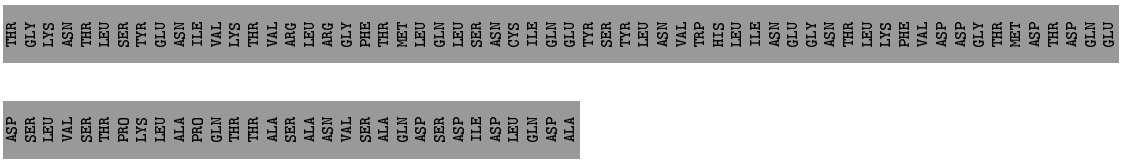


• Molecule 6: DNA replication licensing factor MCM7

Chain 7:

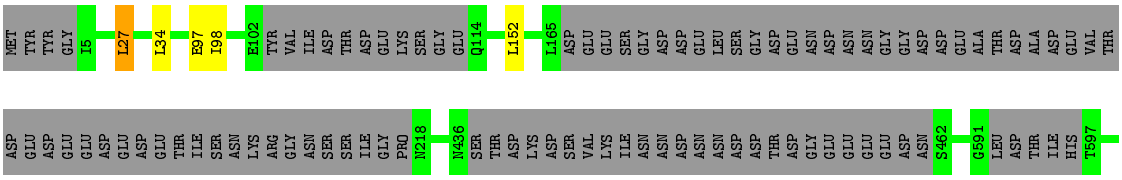






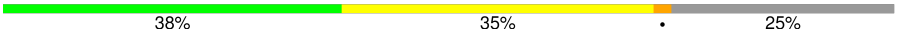
• Molecule 7: Cell division control protein 45

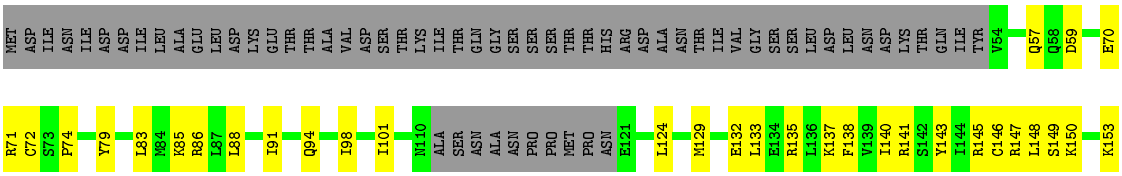
Chain c:

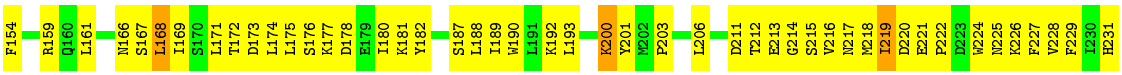


• Molecule 8: DNA replication complex GINS protein SLD5

Chain D:

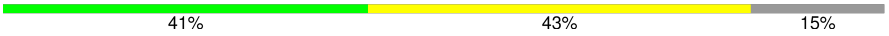


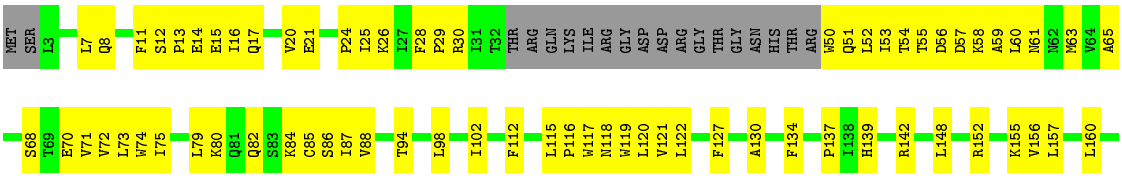


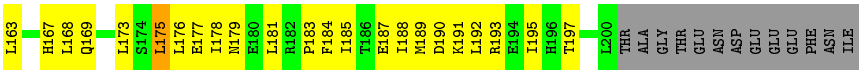


• Molecule 9: DNA replication complex GINS protein PSF2

Chain B:

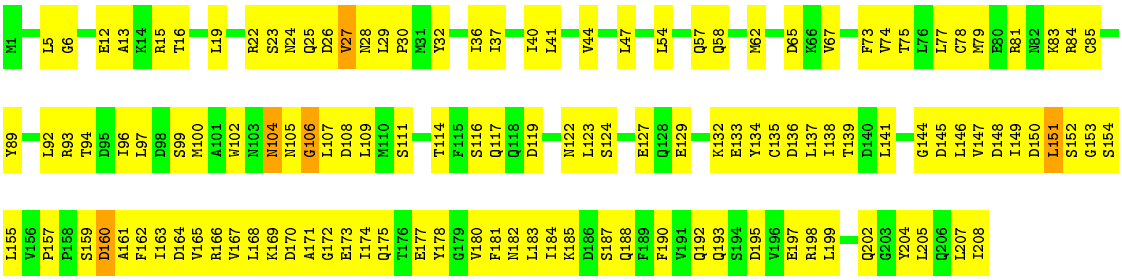






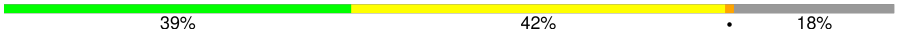
• Molecule 10: DNA replication complex GINS protein PSF1

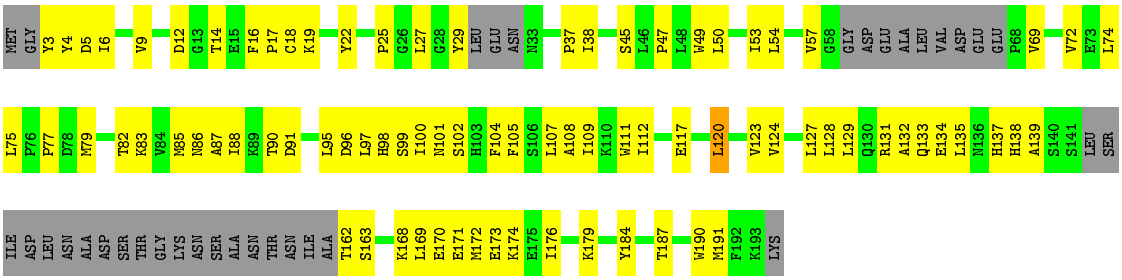
Chain A:



- Molecule 11: DNA replication complex GINS protein PSF3

Chain C:





4 Experimental information (i)

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of images	178530	Depositor
Resolution determination method	FSC 0.143	Depositor
CTF correction method	CTFFIND4	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	50	Depositor
Minimum defocus (nm)	1500	Depositor
Maximum defocus (nm)	3500	Depositor
Magnification	29000	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor

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	#$\|Z\|>2$	RMSZ	#$\|Z\|>2$
1	2	0.41	0/4677	0.62	2/6318 (0.0%)
10	A	0.38	0/1713	0.62	0/2309
11	C	0.46	0/1320	0.64	1/1784 (0.1%)
2	3	0.48	1/4691 (0.0%)	0.66	5/6360 (0.1%)
3	4	0.40	0/4574	0.65	1/6172 (0.0%)
4	5	0.48	1/5242 (0.0%)	0.73	7/7075 (0.1%)
5	6	0.48	1/5289 (0.0%)	0.78	13/7139 (0.2%)
6	7	0.42	0/5228	0.64	1/7062 (0.0%)
7	c	0.39	0/4548	0.63	2/6152 (0.0%)
8	D	0.43	0/1853	0.64	0/2500
9	B	0.43	0/1545	0.69	0/2092
All	All	0.44	3/40680 (0.0%)	0.67	32/54963 (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	2	0	1
10	A	0	4
2	3	0	3
3	4	0	5
5	6	0	6
6	7	0	7
7	c	0	1
8	D	0	2
All	All	0	29

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	3	232	PRO	N-CD	16.13	1.70	1.47

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	6	929	GLU	C-O	7.76	1.38	1.23
4	5	720	ARG	C-N	-7.10	1.17	1.34

The worst 5 of 32 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	6	954	TYR	CB-CG-CD2	-16.55	111.07	121.00
4	5	742	ARG	CA-CB-CG	10.90	137.38	113.40
5	6	954	TYR	CB-CG-CD1	9.43	126.66	121.00
2	3	550	THR	N-CA-C	-9.05	86.57	111.00
4	5	742	ARG	N-CA-C	8.60	134.21	111.00

There are no chirality outliers.

5 of 29 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	2	803	PHE	Peptide
2	3	163	ALA	Peptide
2	3	165	ALA	Peptide
2	3	428	LEU	Peptide
3	4	202	LYS	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	2	4600	0	4642	458	0
2	3	4613	0	4673	583	0
3	4	4516	0	4542	467	0
4	5	5171	0	5232	691	0
5	6	5211	0	5159	513	0
6	7	5148	0	5219	561	0
7	c	4470	0	4491	0	0
8	D	1820	0	1824	160	0
9	B	1513	0	1558	152	0
10	A	1691	0	1687	259	0
11	C	1288	0	1298	108	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	40041	0	40325	3360	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 45.

The worst 5 of 3360 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)
1:2:814:LEU:HD23	4:5:576:HIS:CD2	1.26	1.61
1:2:807:VAL:CB	4:5:572:VAL:HG21	1.31	1.54
3:4:721:ALA:HB2	6:7:664:TYR:CE2	1.42	1.52
3:4:712:VAL:CG1	3:4:715:LYS:HG2	1.39	1.51
1:2:703:HIS:CD2	5:6:804:ILE:HG12	1.44	1.50

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 PDB entries followed by that with respect to all EM entries.

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	2	574/868 (66%)	521 (91%)	44 (8%)	9 (2%)	12 57
2	3	578/971 (60%)	503 (87%)	57 (10%)	18 (3%)	5 43
3	4	547/933 (59%)	475 (87%)	61 (11%)	11 (2%)	9 53
4	5	637/775 (82%)	555 (87%)	61 (10%)	21 (3%)	5 42
5	6	661/1017 (65%)	591 (89%)	58 (9%)	12 (2%)	11 54
6	7	642/845 (76%)	561 (87%)	73 (11%)	8 (1%)	16 62
7	c	543/650 (84%)	496 (91%)	45 (8%)	2 (0%)	39 80
8	D	215/294 (73%)	197 (92%)	16 (7%)	2 (1%)	21 67
9	B	177/213 (83%)	163 (92%)	14 (8%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
10	A	206/208 (99%)	179 (87%)	26 (13%)	1 (0%)	34 77
11	C	151/194 (78%)	144 (95%)	7 (5%)	0	100 100
All	All	4931/6968 (71%)	4385 (89%)	462 (9%)	84 (2%)	16 56

5 of 84 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	2	291	SER
2	3	499	LYS
2	3	517	ASN
2	3	569	HIS
3	4	450	GLN

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 PDB entries followed by that with respect to all EM entries.

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	2	503/770 (65%)	503 (100%)	0	100 100
2	3	511/835 (61%)	511 (100%)	0	100 100
3	4	509/848 (60%)	509 (100%)	0	100 100
4	5	588/688 (86%)	586 (100%)	2 (0%)	94 96
5	6	547/886 (62%)	547 (100%)	0	100 100
6	7	576/753 (76%)	576 (100%)	0	100 100
7	c	498/585 (85%)	495 (99%)	3 (1%)	90 95
8	D	213/279 (76%)	212 (100%)	1 (0%)	92 96
9	B	171/198 (86%)	170 (99%)	1 (1%)	90 95
10	A	192/192 (100%)	191 (100%)	1 (0%)	92 96
11	C	144/173 (83%)	144 (100%)	0	100 100
All	All	4452/6207 (72%)	4444 (100%)	8 (0%)	95 97

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
7	c	34	LEU
10	A	151	LEU
8	D	168	LEU
7	c	27	LEU
7	c	152	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 63 such sidechains are listed below:

Mol	Chain	Res	Type
4	5	259	GLN
5	6	653	HIS
10	A	104	ASN
4	5	411	ASN
4	5	500	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 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.