wwPDB X-ray Structure Validation Summary Report (i)

Jan 2, 2024 – 06:36 pm GMT

PDB ID	:	4UV3
Title	:	Structure of the curli transport lipoprotein CsgG in its membrane- bound conformation
Authors	:	Goyal, P.; Krasteva, P.V.; Gerven, N.V.; Gubellini, F.; Broeck, I.V.D.; Troupiotis-Tsailaki, A.; Jonckheere, W.; Pehau-Arnaudet, G.; Pinkner, J.S.; Chapman, M.R.; Hultgren, S.J.; Howorka, S.; Fronzes, R.; Remaut, H.
Deposited on	:	2014-08-04
Resolution	:	3.59 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 (1)) were used in the production of this report:

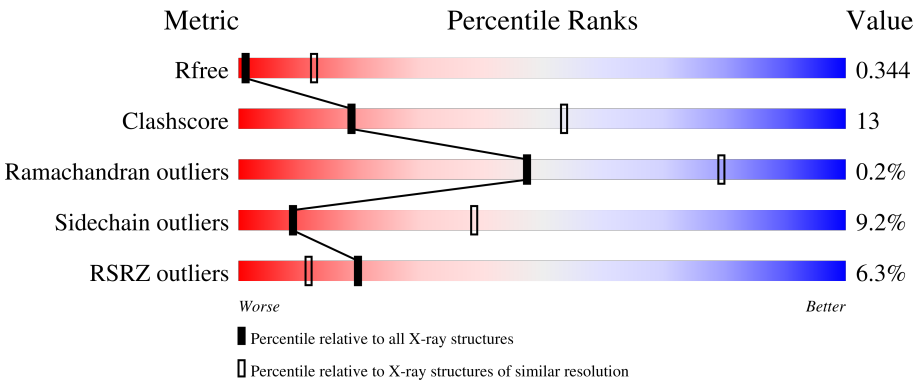
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 3.59 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	1257 (3.70-3.50)
Clashscore	141614	1353 (3.70-3.50)
Ramachandran outliers	138981	1307 (3.70-3.50)
Sidechain outliers	138945	1307 (3.70-3.50)
RSRZ outliers	127900	1161 (3.70-3.50)

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	262	3% 66% 26% • 5%
1	B	262	2% 66% 26% • 5%
1	C	262	8% 66% 27% • 6%
1	D	262	11% 66% 25% • 5%
1	E	262	13% 62% 30% • 6%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	262	6% 61% 29% 6%
1	G	262	4% 62% 28% 6%
1	H	262	2% 66% 27% 5%
1	I	262	6% 63% 29% 6%
1	J	262	2% 62% 30% 5%
1	K	262	5% 63% 29% 6%
1	L	262	2% 66% 25% 6%
1	M	262	2% 66% 25% 6%
1	N	262	8% 66% 27% 6%
1	O	262	10% 65% 26% 6%
1	P	262	12% 66% 26% 6%
1	Q	262	7% 65% 27% 6%
1	R	262	3% 61% 30% 6%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 34255 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 CURLI PRODUCTION ASSEMBLY/TRANSPORT COMPONENT CSGG.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	248	1911	1204	333	367	7	0	0	0
1	B	248	1911	1204	333	367	7	0	0	0
1	C	247	1900	1198	329	366	7	0	0	0
1	D	248	1911	1204	333	367	7	0	0	0
1	E	247	1900	1198	329	366	7	0	0	0
1	F	247	1900	1198	329	366	7	0	0	0
1	G	247	1900	1198	329	366	7	0	0	0
1	H	248	1911	1204	333	367	7	0	0	0
1	I	247	1900	1198	329	366	7	0	0	0
1	J	248	1911	1204	333	367	7	0	0	0
1	K	247	1900	1198	329	366	7	0	0	0
1	L	247	1900	1198	329	366	7	0	0	0
1	M	247	1900	1198	329	366	7	0	0	0
1	N	247	1900	1198	329	366	7	0	0	0
1	O	247	1900	1198	329	366	7	0	0	0
1	P	247	1900	1198	329	366	7	0	0	0

Continued on next page...

Continued from previous page...

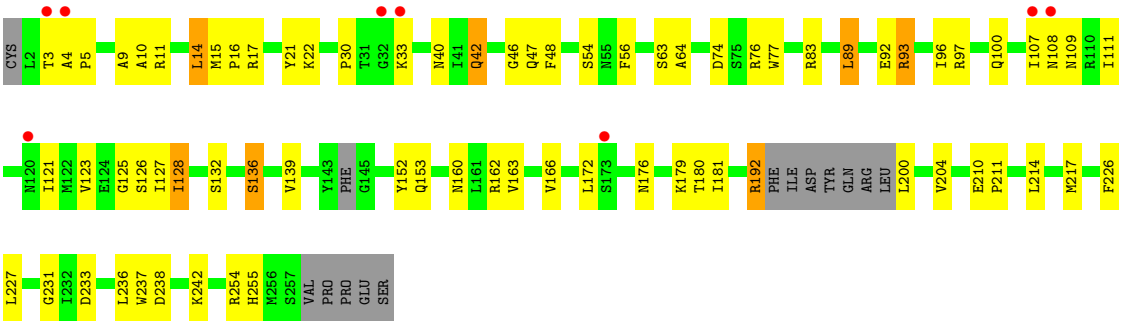
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	Q	247	1900	1198	329	366	7	0	0	0
			Total	C	N	O	S			
1	R	247	1900	1198	329	366	7	0	0	0

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.

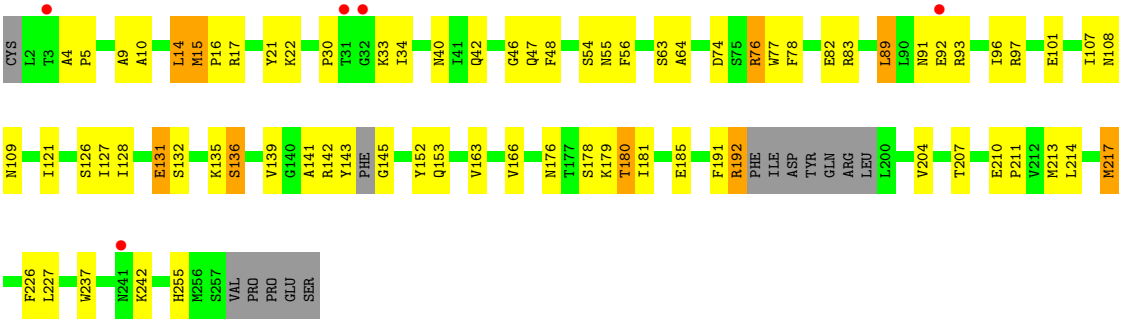
• Molecule 1: CURLI PRODUCTION ASSEMBLY/TRANSPORT COMPONENT CSGG

Chain A:



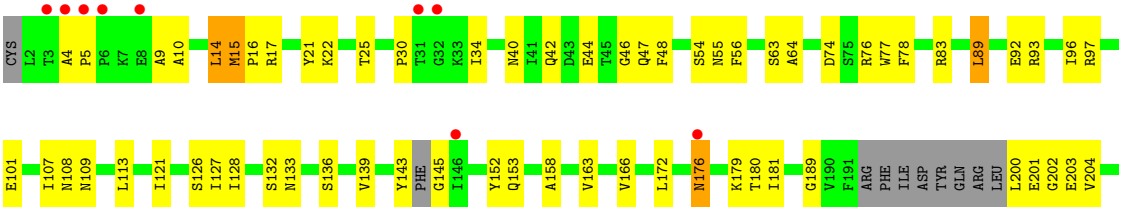
• Molecule 1: CURLI PRODUCTION ASSEMBLY/TRANSPORT COMPONENT CSGG

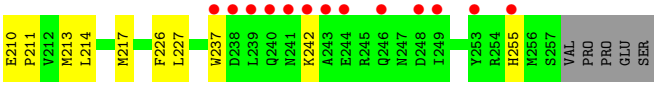
Chain B:



• Molecule 1: CURLI PRODUCTION ASSEMBLY/TRANSPORT COMPONENT CSGG

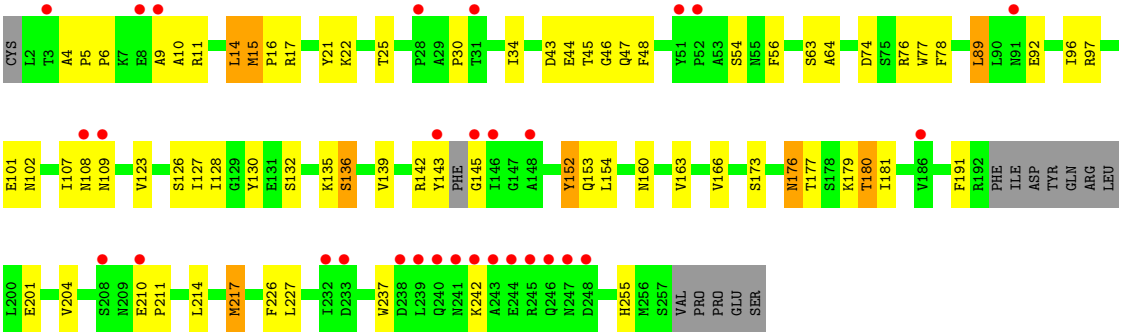
Chain C:





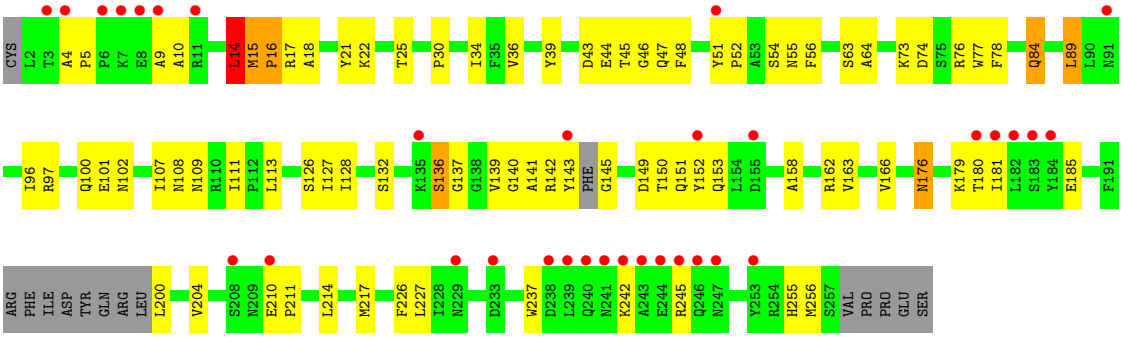
• Molecule 1: CURLI PRODUCTION ASSEMBLY/TRANSPORT COMPONENT CSGG

Chain D:



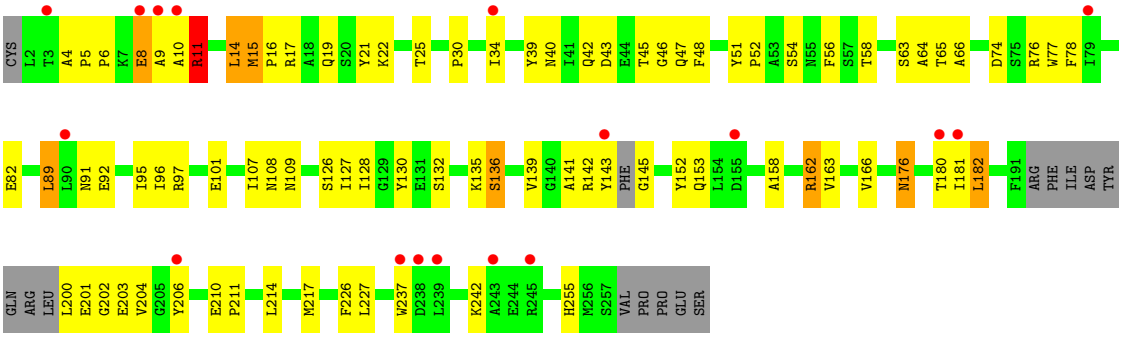
• Molecule 1: CURLI PRODUCTION ASSEMBLY/TRANSPORT COMPONENT CSGG

Chain E:



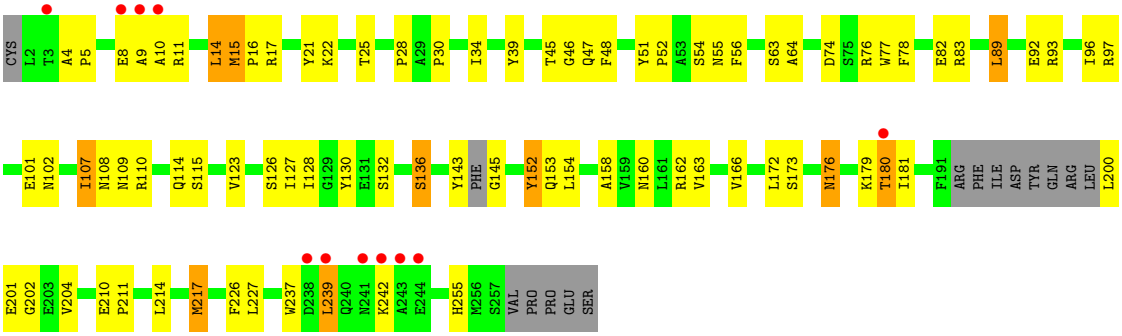
• Molecule 1: CURLI PRODUCTION ASSEMBLY/TRANSPORT COMPONENT CSGG

Chain F:



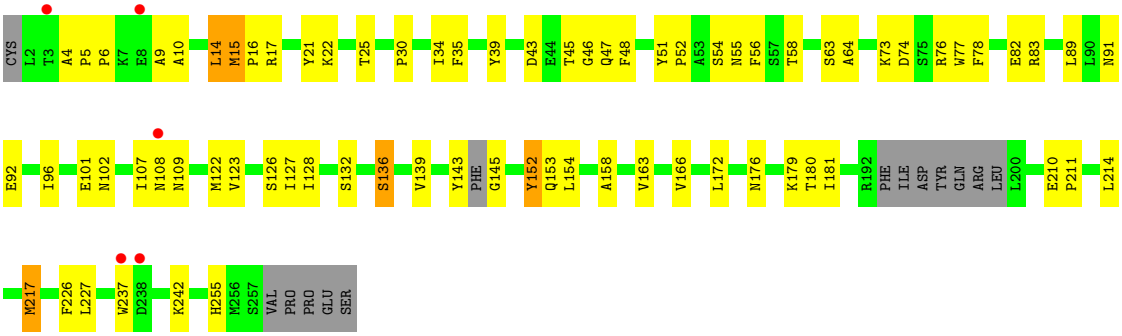
• Molecule 1: CURLI PRODUCTION ASSEMBLY/TRANSPORT COMPONENT CSGG

Chain G:



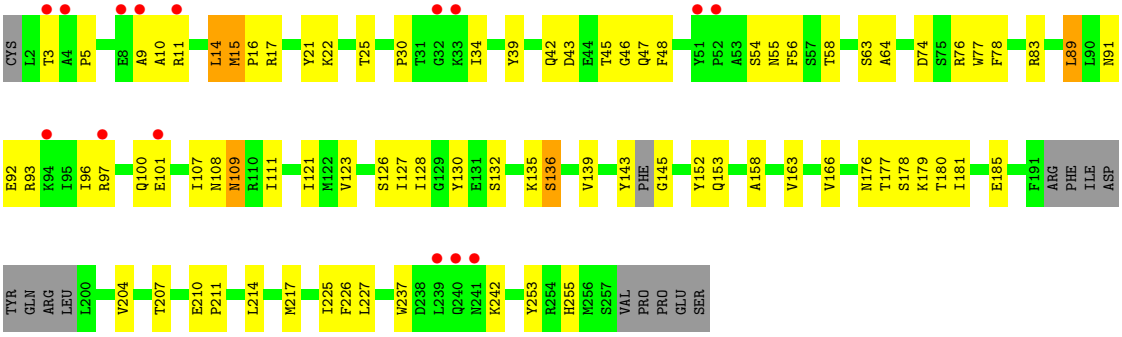
- Molecule 1: CURLI PRODUCTION ASSEMBLY/TRANSPORT COMPONENT CSGG

Chain H:



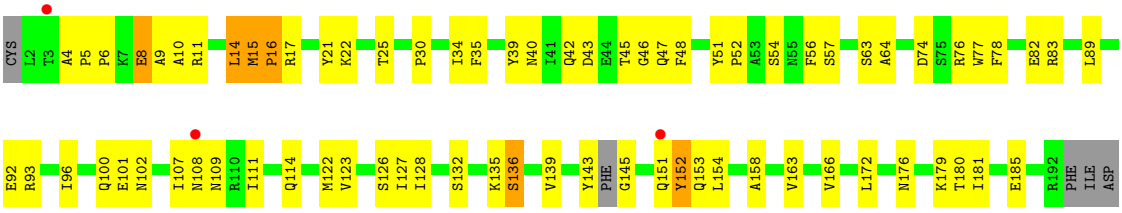
- Molecule 1: CURLI PRODUCTION ASSEMBLY/TRANSPORT COMPONENT CSGG

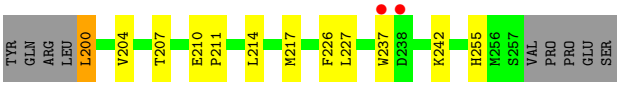
Chain I:



- Molecule 1: CURLI PRODUCTION ASSEMBLY/TRANSPORT COMPONENT CSGG

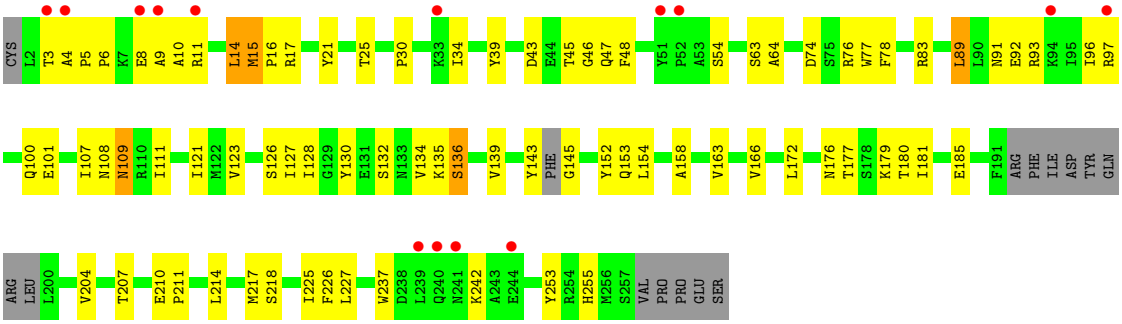
Chain J:





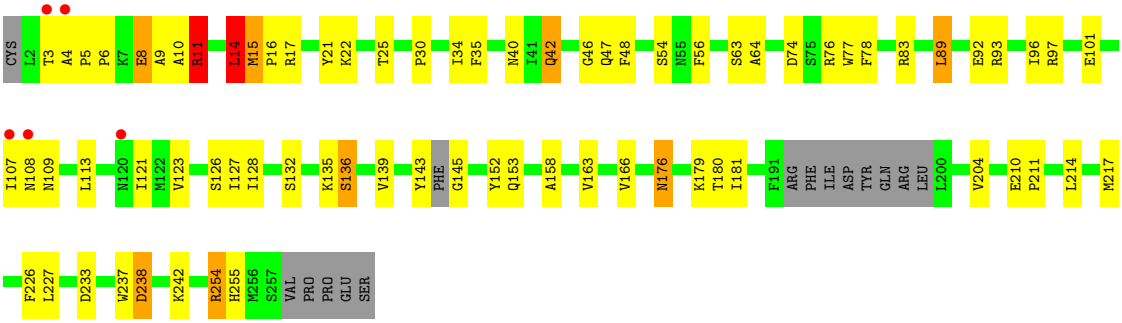
• Molecule 1: CURLI PRODUCTION ASSEMBLY/TRANSPORT COMPONENT CSGG

Chain K:



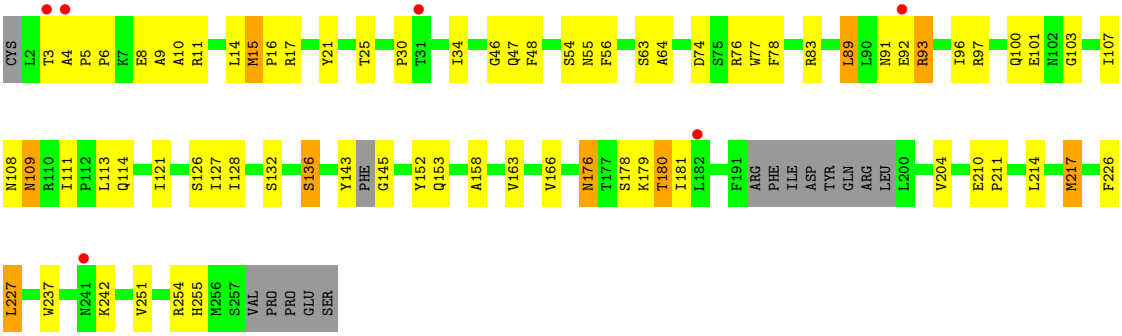
• Molecule 1: CURLI PRODUCTION ASSEMBLY/TRANSPORT COMPONENT CSGG

Chain L:



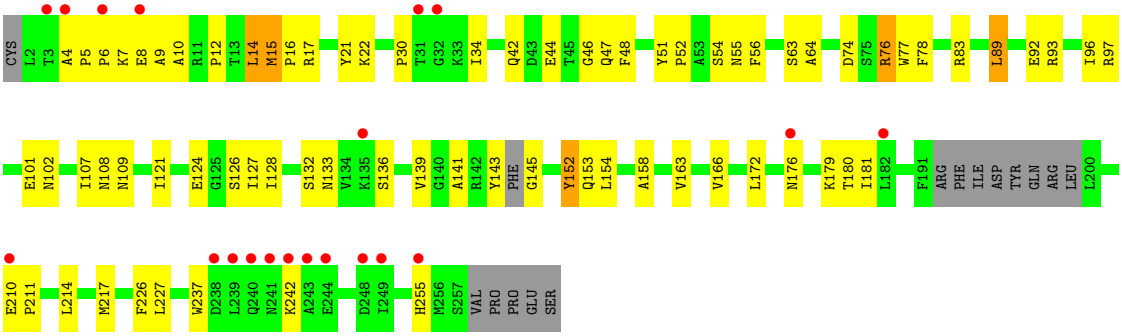
• Molecule 1: CURLI PRODUCTION ASSEMBLY/TRANSPORT COMPONENT CSGG

Chain M:



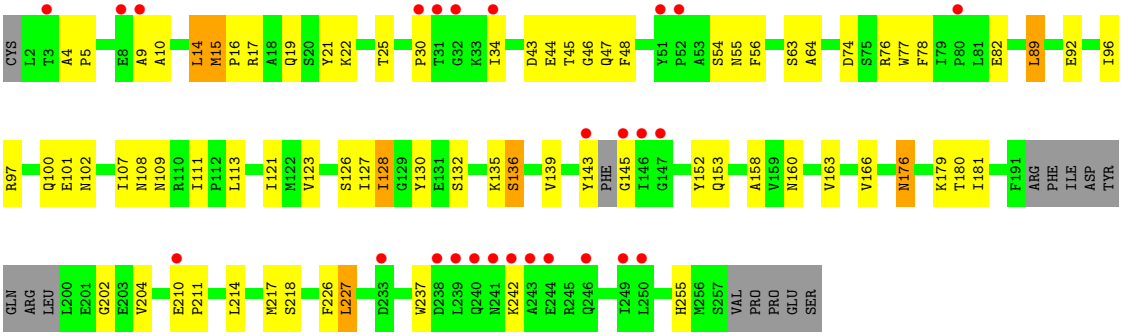
• Molecule 1: CURLI PRODUCTION ASSEMBLY/TRANSPORT COMPONENT CSGG

Chain N:



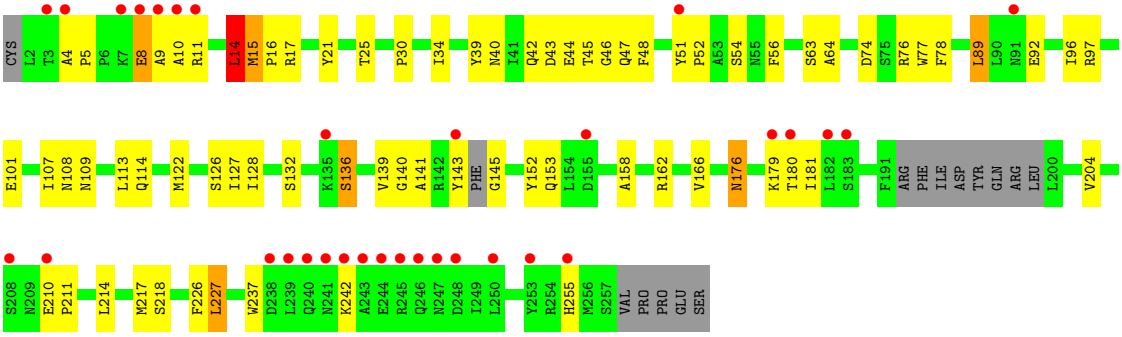
• Molecule 1: CURLI PRODUCTION ASSEMBLY/TRANSPORT COMPONENT CSGG

Chain O:



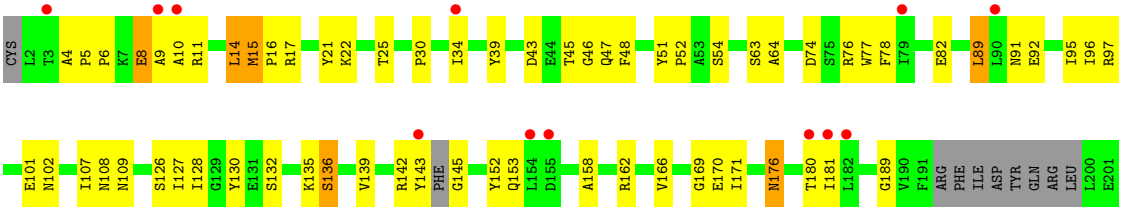
• Molecule 1: CURLI PRODUCTION ASSEMBLY/TRANSPORT COMPONENT CSGG

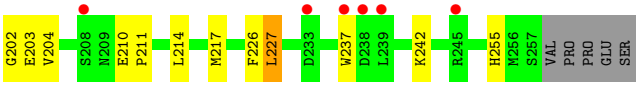
Chain P:



• Molecule 1: CURLI PRODUCTION ASSEMBLY/TRANSPORT COMPONENT CSGG

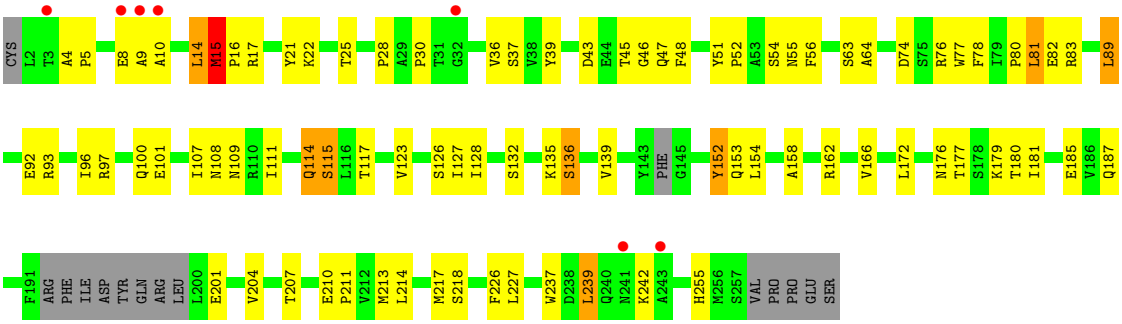
Chain Q:





• Molecule 1: CURLI PRODUCTION ASSEMBLY/TRANSPORT COMPONENT CSGG

Chain R:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	161.93Å 372.85Å 161.97Å 90.00° 92.90° 90.00°	Depositor
Resolution (Å)	30.00 – 3.59 43.85 – 3.59	Depositor EDS
% Data completeness (in resolution range)	91.5 (30.00-3.59) 91.6 (43.85-3.59)	Depositor EDS
R_{merge}	0.16	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.93 (at 3.57Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, R_{free}	0.297 , 0.349 0.296 , 0.344	Depositor DCC
R_{free} test set	5110 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	101.0	Xtriage
Anisotropy	0.225	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 90.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.28$	Xtriage
Estimated twinning fraction	0.029 for h,-k,-l	Xtriage
F_o,F_c correlation	0.81	EDS
Total number of atoms	34255	wwPDB-VP
Average B, all atoms (Å^2)	116.0	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 53.94 % 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 3.9236e-05. 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 (i)

5.1 Standard geometry (i)

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	0.62	0/1941	0.75	1/2633 (0.0%)
1	B	0.57	0/1941	0.74	1/2633 (0.0%)
1	C	0.53	0/1930	0.70	0/2619
1	D	0.52	0/1941	0.68	0/2633
1	E	0.59	0/1930	0.73	0/2619
1	F	0.58	0/1930	0.74	3/2619 (0.1%)
1	G	0.58	0/1930	0.73	0/2619
1	H	0.56	0/1941	0.71	0/2633
1	I	0.55	0/1930	0.70	0/2619
1	J	0.56	0/1941	0.69	0/2633
1	K	0.56	0/1930	0.70	0/2619
1	L	0.62	0/1930	0.77	3/2619 (0.1%)
1	M	0.57	0/1930	0.74	2/2619 (0.1%)
1	N	0.52	0/1930	0.70	0/2619
1	O	0.51	0/1930	0.70	1/2619 (0.0%)
1	P	0.56	0/1930	0.71	0/2619
1	Q	0.57	0/1930	0.71	0/2619
1	R	0.64	2/1930 (0.1%)	0.76	1/2619 (0.0%)
All	All	0.57	2/34795 (0.0%)	0.72	12/47212 (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	E	0	1
1	L	0	1
1	P	0	1
1	R	0	1
All	All	0	4

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	R	115	SER	C-O	-5.42	1.13	1.23
1	R	15	MET	CG-SD	-5.03	1.68	1.81

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	R	81	LEU	C-N-CA	5.73	136.02	121.70
1	O	128	ILE	CB-CA-C	-5.68	100.24	111.60
1	F	11	ARG	NE-CZ-NH1	5.55	123.07	120.30
1	L	11	ARG	NE-CZ-NH1	5.54	123.07	120.30
1	L	238	ASP	CB-CG-OD1	5.53	123.27	118.30

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	E	14	LEU	Peptide
1	L	14	LEU	Peptide
1	P	14	LEU	Peptide
1	R	78	PHE	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	1911	0	1930	70	0
1	B	1911	0	1930	60	0
1	C	1900	0	1917	63	0
1	D	1911	0	1930	67	0
1	E	1900	0	1917	82	0
1	F	1900	0	1917	90	0
1	G	1900	0	1917	86	0
1	H	1911	0	1930	85	0
1	I	1900	0	1917	68	0
1	J	1911	0	1930	90	0
1	K	1900	0	1917	71	0
1	L	1900	0	1917	83	0
1	M	1900	0	1917	68	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	N	1900	0	1917	67	0
1	O	1900	0	1917	69	0
1	P	1900	0	1917	67	0
1	Q	1900	0	1917	67	0
1	R	1900	0	1917	80	0
All	All	34255	0	34571	909	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 13.

The worst 5 of 909 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:F:107:ILE:HD13	1:N:107:ILE:HD13	1.33	1.09
1:E:107:ILE:HD13	1:O:107:ILE:HD13	1.38	1.05
1:E:84:GLN:O	1:E:84:GLN:HG3	1.57	0.99
1:D:107:ILE:HD13	1:P:107:ILE:HD13	1.44	0.99
1:F:182:LEU:HD22	1:G:9:ALA:HB1	1.43	0.99

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	242/262 (92%)	225 (93%)	16 (7%)	1 (0%)	34 71
1	B	242/262 (92%)	227 (94%)	15 (6%)	0	100 100
1	C	241/262 (92%)	225 (93%)	15 (6%)	1 (0%)	34 71
1	D	242/262 (92%)	227 (94%)	14 (6%)	1 (0%)	34 71
1	E	241/262 (92%)	226 (94%)	13 (5%)	2 (1%)	19 59

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	F	241/262 (92%)	228 (95%)	13 (5%)	0	100	100
1	G	241/262 (92%)	227 (94%)	14 (6%)	0	100	100
1	H	242/262 (92%)	226 (93%)	16 (7%)	0	100	100
1	I	241/262 (92%)	229 (95%)	12 (5%)	0	100	100
1	J	242/262 (92%)	225 (93%)	16 (7%)	1 (0%)	34	71
1	K	241/262 (92%)	227 (94%)	14 (6%)	0	100	100
1	L	241/262 (92%)	226 (94%)	15 (6%)	0	100	100
1	M	241/262 (92%)	226 (94%)	15 (6%)	0	100	100
1	N	241/262 (92%)	224 (93%)	16 (7%)	1 (0%)	34	71
1	O	241/262 (92%)	226 (94%)	14 (6%)	1 (0%)	34	71
1	P	241/262 (92%)	227 (94%)	13 (5%)	1 (0%)	34	71
1	Q	241/262 (92%)	228 (95%)	13 (5%)	0	100	100
1	R	241/262 (92%)	225 (93%)	16 (7%)	0	100	100
All	All	4343/4716 (92%)	4074 (94%)	260 (6%)	9 (0%)	47	79

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	44	GLU
1	O	44	GLU
1	P	44	GLU
1	C	44	GLU
1	D	44	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	209/223 (94%)	189 (90%)	20 (10%)	8	37
1	B	209/223 (94%)	188 (90%)	21 (10%)	7	35
1	C	208/223 (93%)	190 (91%)	18 (9%)	10	41

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	D	209/223 (94%)	191 (91%)	18 (9%)	10	41
1	E	208/223 (93%)	190 (91%)	18 (9%)	10	41
1	F	208/223 (93%)	190 (91%)	18 (9%)	10	41
1	G	208/223 (93%)	187 (90%)	21 (10%)	7	34
1	H	209/223 (94%)	191 (91%)	18 (9%)	10	41
1	I	208/223 (93%)	189 (91%)	19 (9%)	9	39
1	J	209/223 (94%)	188 (90%)	21 (10%)	7	35
1	K	208/223 (93%)	190 (91%)	18 (9%)	10	41
1	L	208/223 (93%)	186 (89%)	22 (11%)	6	32
1	M	208/223 (93%)	191 (92%)	17 (8%)	11	42
1	N	208/223 (93%)	189 (91%)	19 (9%)	9	39
1	O	208/223 (93%)	190 (91%)	18 (9%)	10	41
1	P	208/223 (93%)	189 (91%)	19 (9%)	9	39
1	Q	208/223 (93%)	189 (91%)	19 (9%)	9	39
1	R	208/223 (93%)	187 (90%)	21 (10%)	7	34
All	All	3749/4014 (93%)	3404 (91%)	345 (9%)	9	39

5 of 345 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	L	255	HIS
1	P	8	GLU
1	M	126	SER
1	N	136	SER
1	P	217	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 31 such sidechains are listed below:

Mol	Chain	Res	Type
1	G	176	ASN
1	P	151	GLN
1	J	40	ASN
1	Q	160	ASN
1	N	153	GLN

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 (i)

6.1 Protein, DNA and RNA chains (i)

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, 95th 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	248/262 (94%)	-0.19	8 (3%)	47	32	38, 94, 165, 213	0
1	B	248/262 (94%)	-0.06	5 (2%)	65	49	45, 99, 162, 200	0
1	C	247/262 (94%)	0.36	22 (8%)	9	5	49, 119, 181, 226	0
1	D	248/262 (94%)	0.55	30 (12%)	4	3	58, 120, 186, 224	0
1	E	247/262 (94%)	0.58	33 (13%)	3	2	52, 118, 183, 230	0
1	F	247/262 (94%)	0.24	17 (6%)	16	10	53, 117, 184, 233	0
1	G	247/262 (94%)	0.18	11 (4%)	33	21	53, 117, 174, 236	0
1	H	248/262 (94%)	0.11	5 (2%)	65	49	57, 112, 174, 249	0
1	I	247/262 (94%)	0.07	15 (6%)	21	12	43, 105, 167, 258	0
1	J	248/262 (94%)	0.10	5 (2%)	65	49	58, 112, 184, 238	0
1	K	247/262 (94%)	0.02	14 (5%)	23	14	38, 107, 168, 229	0
1	L	247/262 (94%)	-0.20	5 (2%)	65	49	29, 92, 155, 226	0
1	M	247/262 (94%)	-0.01	6 (2%)	59	42	41, 102, 158, 205	0
1	N	247/262 (94%)	0.32	20 (8%)	12	7	42, 115, 173, 212	0
1	O	247/262 (94%)	0.52	26 (10%)	6	3	57, 123, 184, 229	0
1	P	247/262 (94%)	0.54	32 (12%)	3	2	51, 120, 188, 248	0
1	Q	247/262 (94%)	0.25	18 (7%)	15	9	50, 121, 184, 240	0
1	R	247/262 (94%)	0.10	7 (2%)	53	37	50, 116, 176, 204	0
All	All	4451/4716 (94%)	0.19	279 (6%)	20	11	29, 111, 179, 258	0

The worst 5 of 279 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	243	ALA	6.8
1	F	9	ALA	6.7
1	O	143	TYR	6.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	D	143	TYR	6.2
1	D	239	LEU	6.1

6.2 Non-standard residues in protein, DNA, RNA chains (i)

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.