wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 09:07 PM GMT

PDB ID	:	4UV2
Title	:	Structure of the curli transport lipoprotein CsgG in a non-lipidated, pre-pore 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	:	2.80 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
http://wwpdb.org/validation/2016/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:

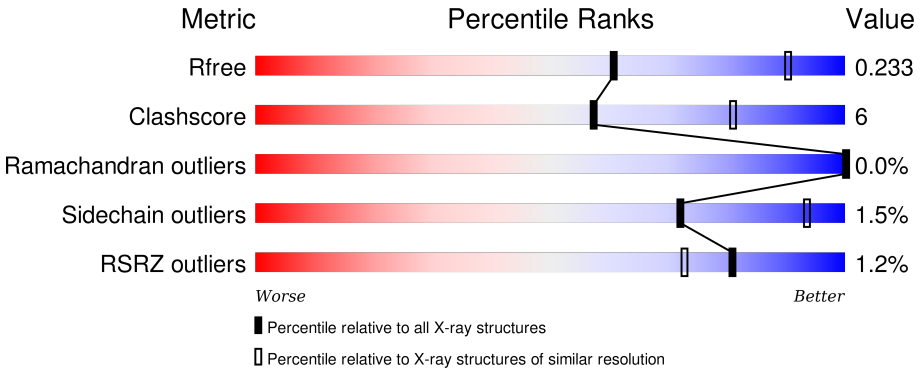
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 2.80 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.



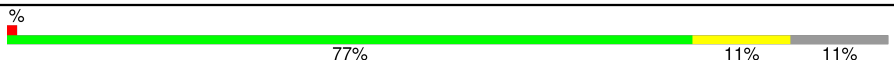
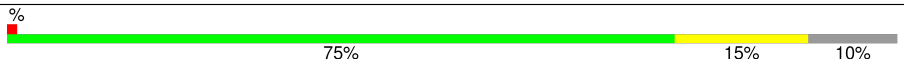
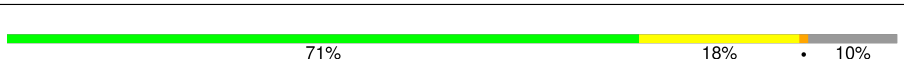
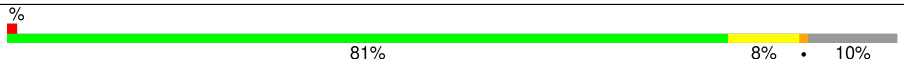
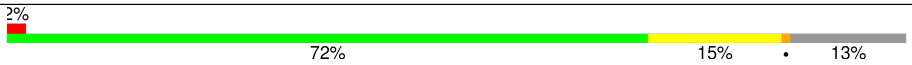
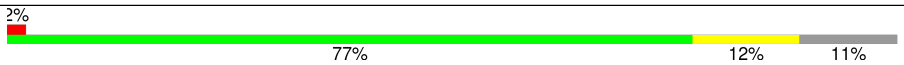
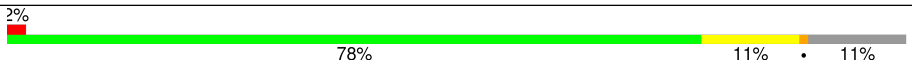
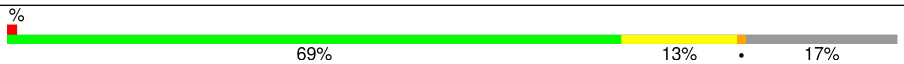
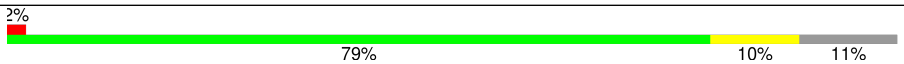
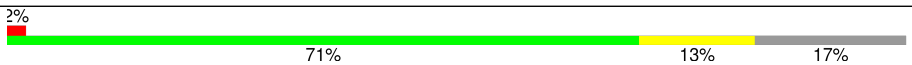
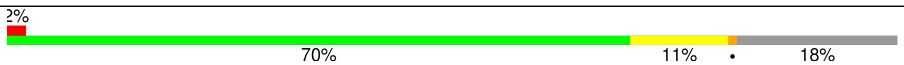
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	2393 (2.80-2.80)
Clashscore	102246	2827 (2.80-2.80)
Ramachandran outliers	100387	2782 (2.80-2.80)
Sidechain outliers	100360	2784 (2.80-2.80)
RSRZ outliers	91569	2404 (2.80-2.80)

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. 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	% 77% 12% 11%
1	B	262	% 80% 10% 10%
1	C	262	79% 11% 11%
1	D	262	74% 15% 10%
1	E	262	81% 11% • 8%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	262	
1	G	262	
1	H	262	
1	I	262	
1	J	262	
1	K	262	
1	L	262	
1	M	262	
1	N	262	
1	O	262	
1	P	262	

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 28853 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 TRANSPORT COMPONENT CSGG.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	234	Total	C	N	O	S	Se	95	0	0
			1817	1151	312	348	1	5			
1	B	236	Total	C	N	O	S	Se	75	0	0
			1835	1162	317	350	1	5			
1	C	234	Total	C	N	O	S	Se	41	0	0
			1817	1151	312	348	1	5			
1	D	235	Total	C	N	O	S	Se	68	0	0
			1828	1157	316	349	1	5			
1	E	241	Total	C	N	O	S	Se	91	0	0
			1879	1187	326	360	1	5			
1	F	234	Total	C	N	O	S	Se	115	0	0
			1821	1152	315	348	1	5			
1	G	235	Total	C	N	O	S	Se	121	0	0
			1830	1157	317	350	1	5			
1	H	236	Total	C	N	O	S	Se	72	0	0
			1835	1162	317	350	1	5			
1	I	235	Total	C	N	O	S	Se	117	0	0
			1828	1157	316	349	1	5			
1	J	228	Total	C	N	O	S	Se	110	0	0
			1785	1131	309	339	1	5			
1	K	234	Total	C	N	O	S	Se	119	0	0
			1821	1152	315	348	1	5			
1	L	234	Total	C	N	O	S	Se	119	0	0
			1821	1152	315	348	1	5			
1	M	217	Total	C	N	O	S	Se	80	0	0
			1705	1079	294	326	1	5			
1	N	233	Total	C	N	O	S	Se	162	0	0
			1814	1147	314	347	1	5			
1	O	218	Total	C	N	O	S	Se	92	0	0
			1714	1085	296	327	1	5			
1	P	216	Total	C	N	O	S	Se	99	0	0
			1703	1080	293	324	1	5			

There are 16 discrepancies between the modelled and reference sequences:

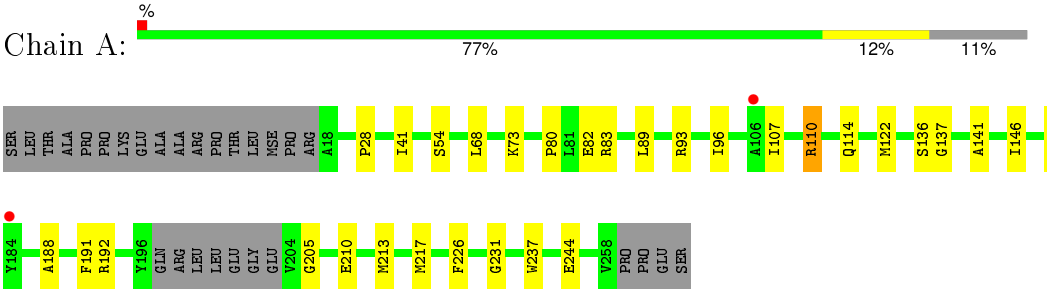
Chain	Residue	Modelled	Actual	Comment	Reference
A	1	SER	-	EXPRESSION TAG	UNP P0AEA2
B	1	SER	-	EXPRESSION TAG	UNP P0AEA2
C	1	SER	-	EXPRESSION TAG	UNP P0AEA2
D	1	SER	-	EXPRESSION TAG	UNP P0AEA2
E	1	SER	-	EXPRESSION TAG	UNP P0AEA2
F	1	SER	-	EXPRESSION TAG	UNP P0AEA2
G	1	SER	-	EXPRESSION TAG	UNP P0AEA2
H	1	SER	-	EXPRESSION TAG	UNP P0AEA2
I	1	SER	-	EXPRESSION TAG	UNP P0AEA2
J	1	SER	-	EXPRESSION TAG	UNP P0AEA2
K	1	SER	-	EXPRESSION TAG	UNP P0AEA2
L	1	SER	-	EXPRESSION TAG	UNP P0AEA2
M	1	SER	-	EXPRESSION TAG	UNP P0AEA2
N	1	SER	-	EXPRESSION TAG	UNP P0AEA2
O	1	SER	-	EXPRESSION TAG	UNP P0AEA2
P	1	SER	-	EXPRESSION TAG	UNP P0AEA2

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 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.

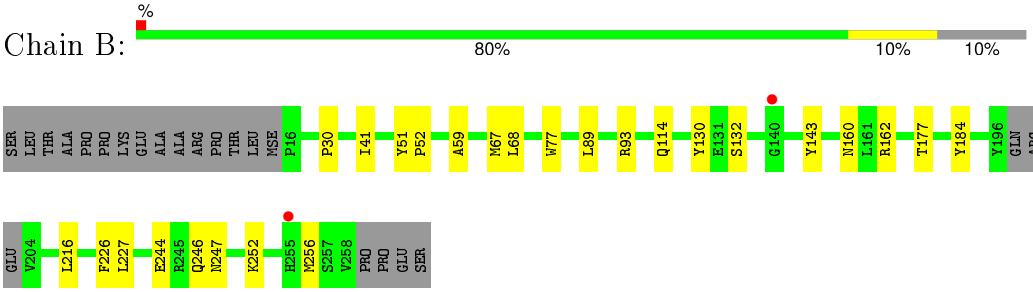
• Molecule 1: CURLI PRODUCTION TRANSPORT COMPONENT CSGG

Chain A:



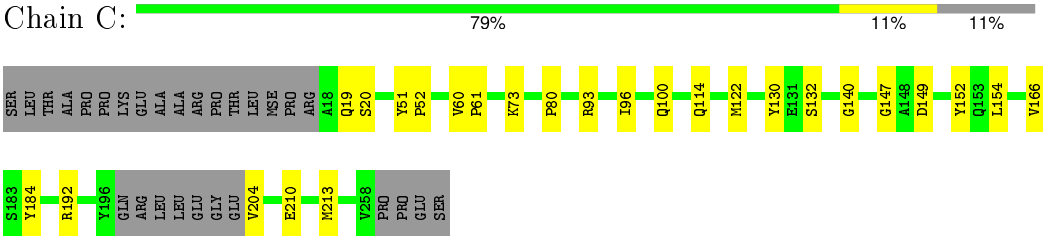
• Molecule 1: CURLI PRODUCTION TRANSPORT COMPONENT CSGG

Chain B:



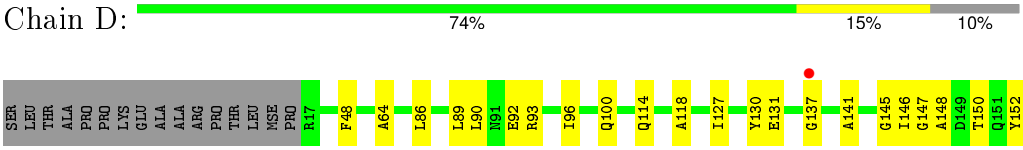
• Molecule 1: CURLI PRODUCTION TRANSPORT COMPONENT CSGG

Chain C:



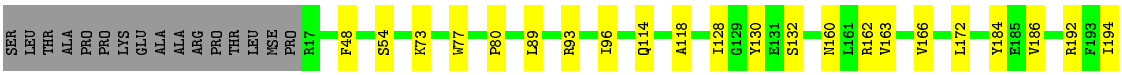
• Molecule 1: CURLI PRODUCTION TRANSPORT COMPONENT CSGG

Chain D:



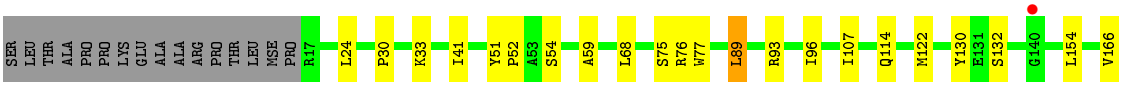
• Molecule 1: CURLI PRODUCTION TRANSPORT COMPONENT CSGG

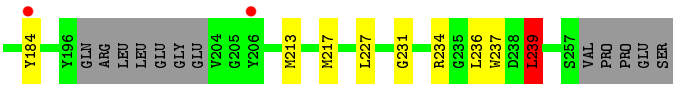
Chain E:



• Molecule 1: CURLI PRODUCTION TRANSPORT COMPONENT CSGG

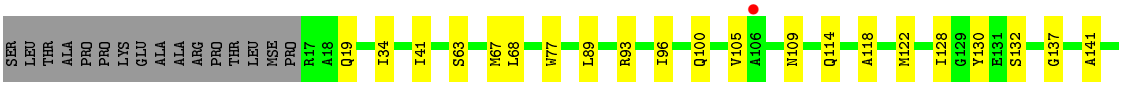
Chain F:

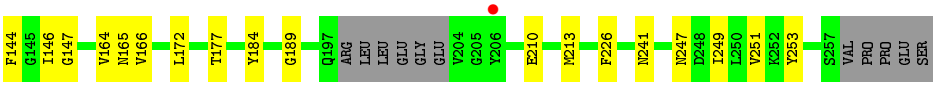




• Molecule 1: CURLI PRODUCTION TRANSPORT COMPONENT CSGG

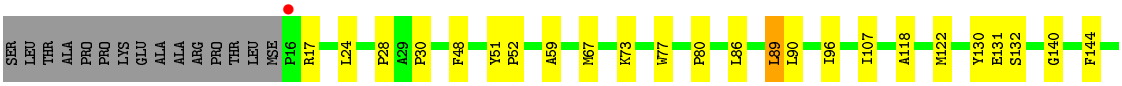
Chain G:





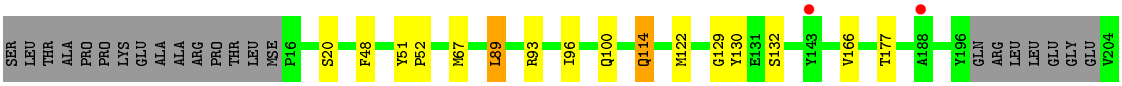
• Molecule 1: CURLI PRODUCTION TRANSPORT COMPONENT CSGG

Chain H:



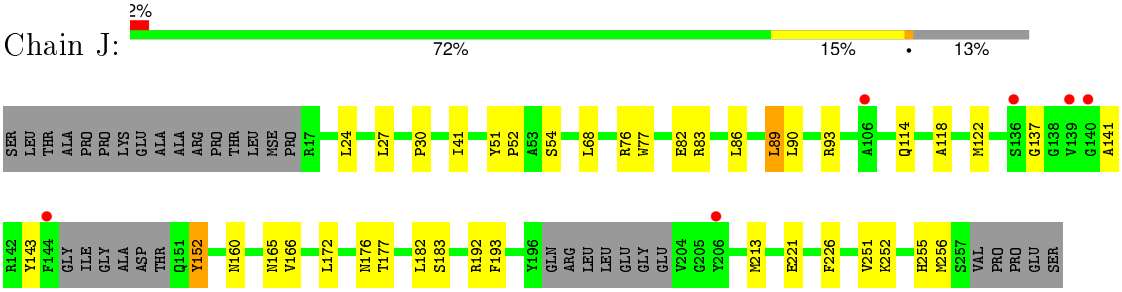
• Molecule 1: CURLI PRODUCTION TRANSPORT COMPONENT CSGG

Chain I:



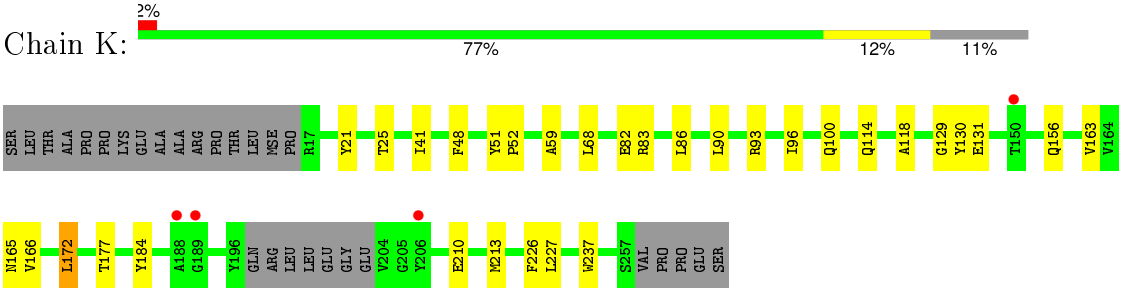
• Molecule 1: CURLI PRODUCTION TRANSPORT COMPONENT CSGG

Chain J:



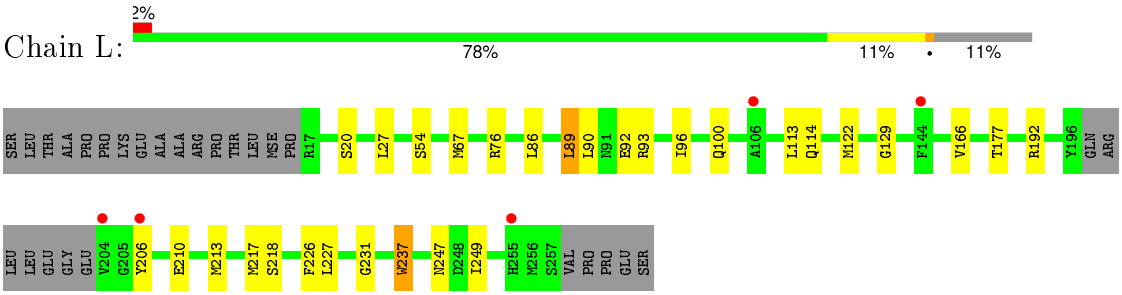
• Molecule 1: CURLI PRODUCTION TRANSPORT COMPONENT CSGG

Chain K:



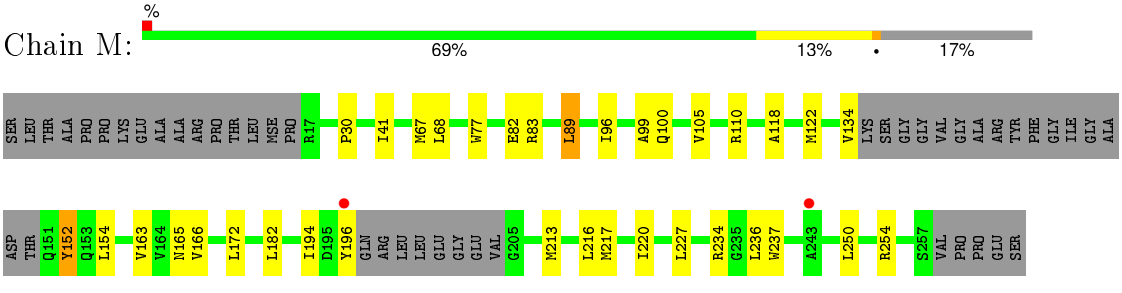
• Molecule 1: CURLI PRODUCTION TRANSPORT COMPONENT CSGG

Chain L:



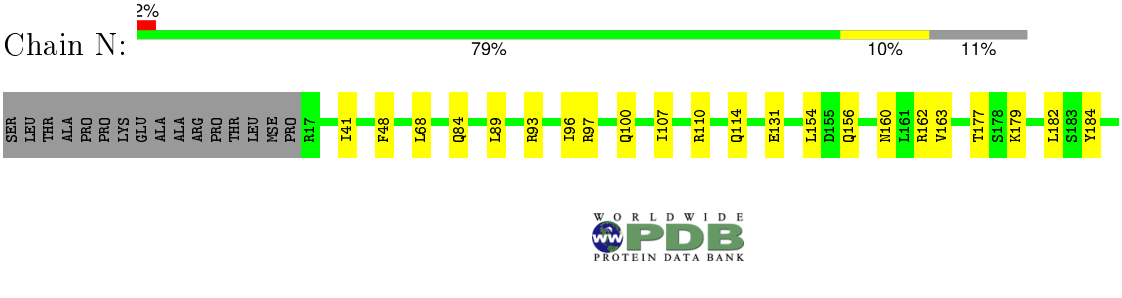
• Molecule 1: CURLI PRODUCTION TRANSPORT COMPONENT CSGG

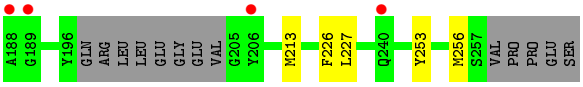
Chain M:



• Molecule 1: CURLI PRODUCTION TRANSPORT COMPONENT CSGG

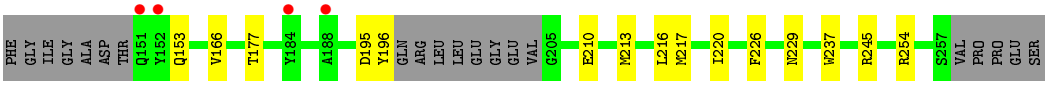
Chain N:





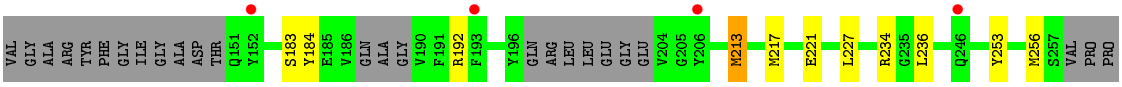
• Molecule 1: CURLI PRODUCTION TRANSPORT COMPONENT CSGG

Chain O:



• Molecule 1: CURLI PRODUCTION TRANSPORT COMPONENT CSGG

Chain P:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	101.33Å 103.60Å 141.74Å 111.33° 90.55° 118.21°	Depositor
Resolution (Å)	29.77 – 2.80 29.76 – 2.80	Depositor EDS
% Data completeness (in resolution range)	99.0 (29.77-2.80) 83.3 (29.76-2.80)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.70 (at 2.80Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, R_{free}	0.188 , 0.234 0.186 , 0.233	Depositor DCC
R_{free} test set	5618 reflections (5.00%)	DCC
Wilson B-factor (Å^2)	46.8	Xtriage
Anisotropy	0.197	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 50.8	EDS
Estimated twinning fraction	0.000 for k,-h-k,h+k+l 0.000 for -h-k,h,k+l 0.002 for h,-h-k,-l 0.019 for -h-k,k,-k-l 0.004 for k,h,-h-k-l	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Outliers	0 of 112419 reflections	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	28853	wwPDB-VP
Average B, all atoms (Å^2)	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.11% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.50	0/1842	0.64	0/2491
1	B	0.51	0/1861	0.65	0/2516
1	C	0.49	0/1842	0.66	0/2491
1	D	0.50	0/1853	0.69	0/2505
1	E	0.48	0/1905	0.62	0/2575
1	F	0.47	0/1846	0.67	1/2495 (0.0%)
1	G	0.50	0/1855	0.71	0/2507
1	H	0.51	0/1861	0.65	0/2516
1	I	0.47	0/1854	0.65	0/2506
1	J	0.47	0/1809	0.66	0/2443
1	K	0.49	0/1846	0.68	0/2495
1	L	0.44	0/1846	0.65	0/2495
1	M	0.49	0/1727	0.66	0/2334
1	N	0.46	0/1839	0.64	0/2485
1	O	0.47	0/1736	0.69	1/2345 (0.0%)
1	P	0.49	0/1724	0.67	0/2328
All	All	0.48	0/29246	0.66	2/39527 (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	O	195	ASP	CB-CG-OD1	5.45	123.21	118.30
1	F	239	LEU	CA-CB-CG	5.05	126.91	115.30

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	1817	0	1823	23	0
1	B	1835	0	1844	18	0
1	C	1817	0	1823	18	0
1	D	1828	0	1836	27	0
1	E	1879	0	1886	19	0
1	F	1821	0	1827	23	0
1	G	1830	0	1835	29	0
1	H	1835	0	1844	38	0
1	I	1828	0	1835	19	0
1	J	1785	0	1793	23	0
1	K	1821	0	1827	22	0
1	L	1821	0	1827	23	0
1	M	1705	0	1712	27	0
1	N	1814	0	1818	23	0
1	O	1714	0	1725	24	0
1	P	1703	0	1717	22	0
All	All	28853	0	28972	326	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 6.

The worst 5 of 326 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:H:28:PRO:HB3	1:H:244:GLU:HG2	1.49	0.95
1:A:122:MSE:HE2	1:A:166:VAL:HG22	1.52	0.90
1:C:122:MSE:HE1	1:D:89:LEU:HA	1.53	0.90
1:O:67:MSE:HE2	1:O:220:ILE:HD12	1.52	0.88
1:M:67:MSE:HE2	1:M:220:ILE:HD12	1.60	0.84

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	230/262 (88%)	228 (99%)	2 (1%)	0	100	100
1	B	232/262 (88%)	229 (99%)	3 (1%)	0	100	100
1	C	230/262 (88%)	228 (99%)	2 (1%)	0	100	100
1	D	231/262 (88%)	227 (98%)	4 (2%)	0	100	100
1	E	239/262 (91%)	234 (98%)	4 (2%)	1 (0%)	39	74
1	F	230/262 (88%)	226 (98%)	4 (2%)	0	100	100
1	G	231/262 (88%)	227 (98%)	4 (2%)	0	100	100
1	H	232/262 (88%)	229 (99%)	3 (1%)	0	100	100
1	I	231/262 (88%)	228 (99%)	3 (1%)	0	100	100
1	J	222/262 (85%)	218 (98%)	4 (2%)	0	100	100
1	K	230/262 (88%)	224 (97%)	6 (3%)	0	100	100
1	L	230/262 (88%)	225 (98%)	5 (2%)	0	100	100
1	M	211/262 (80%)	207 (98%)	4 (2%)	0	100	100
1	N	229/262 (87%)	224 (98%)	5 (2%)	0	100	100
1	O	212/262 (81%)	208 (98%)	4 (2%)	0	100	100
1	P	208/262 (79%)	207 (100%)	1 (0%)	0	100	100
All	All	3628/4192 (86%)	3569 (98%)	58 (2%)	1 (0%)	100	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	196	TYR

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	199/217 (92%)	196 (98%)	3 (2%)	72	93

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	B	201/217 (93%)	198 (98%)	3 (2%)	72 93
1	C	199/217 (92%)	197 (99%)	2 (1%)	82 96
1	D	200/217 (92%)	196 (98%)	4 (2%)	63 90
1	E	205/217 (94%)	201 (98%)	4 (2%)	63 90
1	F	199/217 (92%)	195 (98%)	4 (2%)	63 90
1	G	200/217 (92%)	197 (98%)	3 (2%)	72 93
1	H	201/217 (93%)	194 (96%)	7 (4%)	43 77
1	I	200/217 (92%)	198 (99%)	2 (1%)	82 96
1	J	196/217 (90%)	193 (98%)	3 (2%)	72 93
1	K	199/217 (92%)	196 (98%)	3 (2%)	72 93
1	L	199/217 (92%)	197 (99%)	2 (1%)	82 96
1	M	189/217 (87%)	185 (98%)	4 (2%)	61 90
1	N	198/217 (91%)	197 (100%)	1 (0%)	92 98
1	O	190/217 (88%)	189 (100%)	1 (0%)	92 98
1	P	190/217 (88%)	188 (99%)	2 (1%)	80 95
All	All	3165/3472 (91%)	3117 (98%)	48 (2%)	72 93

5 of 48 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	109	ASN
1	H	172	LEU
1	N	184	TYR
1	G	241	ASN
1	H	89	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	209	ASN
1	D	153	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.

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	229/262 (87%)	-0.57	2 (0%) 85 79	25, 50, 102, 128	26 (11%)
1	B	231/262 (88%)	-0.60	2 (0%) 85 79	26, 51, 88, 101	22 (9%)
1	C	229/262 (87%)	-0.72	0 100 100	23, 50, 83, 103	15 (6%)
1	D	230/262 (87%)	-0.59	1 (0%) 93 90	24, 50, 99, 138	22 (9%)
1	E	236/262 (90%)	-0.67	0 100 100	26, 53, 85, 104	30 (12%)
1	F	229/262 (87%)	-0.54	3 (1%) 79 71	27, 54, 101, 132	36 (15%)
1	G	230/262 (87%)	-0.54	2 (0%) 85 79	27, 50, 92, 123	36 (15%)
1	H	231/262 (88%)	-0.69	1 (0%) 93 90	26, 49, 78, 95	29 (12%)
1	I	230/262 (87%)	-0.76	2 (0%) 85 79	26, 53, 86, 106	44 (19%)
1	J	223/262 (85%)	-0.51	6 (2%) 58 45	27, 52, 101, 118	33 (14%)
1	K	229/262 (87%)	-0.59	4 (1%) 73 63	26, 54, 99, 113	36 (15%)
1	L	229/262 (87%)	-0.52	5 (2%) 65 54	28, 59, 103, 116	35 (15%)
1	M	212/262 (80%)	-0.67	2 (0%) 85 79	30, 51, 83, 101	29 (13%)
1	N	228/262 (87%)	-0.53	4 (1%) 71 61	31, 56, 99, 123	53 (23%)
1	O	213/262 (81%)	-0.59	4 (1%) 70 59	31, 52, 91, 110	30 (14%)
1	P	211/262 (80%)	-0.50	4 (1%) 70 59	27, 57, 96, 112	29 (13%)
All	All	3620/4192 (86%)	-0.60	42 (1%) 81 73	23, 52, 93, 138	505 (13%)

The worst 5 of 42 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	J	139	VAL	3.7
1	I	188	ALA	3.6
1	J	136	SER	3.5
1	N	188	ALA	3.4
1	A	184	TYR	3.4

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 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.