wwPDB X-ray Structure Validation Summary Report (i)

Apr 30, 2024 – 08:10 pm BST

PDB ID	:	4B9G
Title	:	Structure of CssB subunit complemented with donor strand from CssA subunit of enterotoxigenic Escherichia coli colonization factor CS6
Authors	:	Roy, S.P.; Rahman, M.M.; Yu, X.D.; Tuittila, M.; Knight, S.D.; Zavialov, A.V.
Deposited on	:	2012-09-04
Resolution	:	1.04 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:

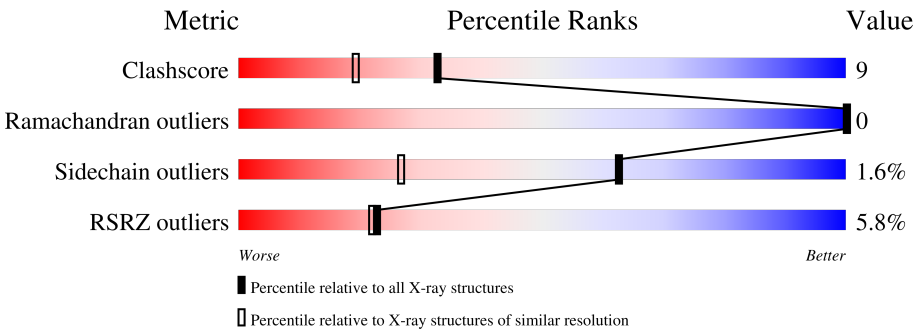
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 1.04 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))
Clashscore	141614	1677 (1.10-0.98)
Ramachandran outliers	138981	1591 (1.10-0.98)
Sidechain outliers	138945	1589 (1.10-0.98)
RSRZ outliers	127900	1557 (1.10-0.98)

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	161	4% 81% 9% • 9%
1	B	161	6% 78% 10% • 9%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 4844 atoms, of which 2152 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 CS6 FIMBRIAL SUBUNIT B, CS6 FIMBRIAL SUBUNIT A.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	146	2189	695	1076	184	231	3	0	0	0
1	B	146	2189	695	1076	184	231	3	0	0	0

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	expression tag	UNP P53510
A	2	ASN	-	expression tag	UNP P53510
A	3	TRP	-	expression tag	UNP P53510
A	142	ASP	-	linker	UNP P53510
A	143	ASN	-	linker	UNP P53510
A	144	LYS	-	linker	UNP P53510
A	145	GLN	-	linker	UNP P53510
B	1	GLY	-	expression tag	UNP P53510
B	2	ASN	-	expression tag	UNP P53510
B	3	TRP	-	expression tag	UNP P53510
B	142	ASP	-	linker	UNP P53510
B	143	ASN	-	linker	UNP P53510
B	144	LYS	-	linker	UNP P53510
B	145	GLN	-	linker	UNP P53510

- Molecule 2 is water.

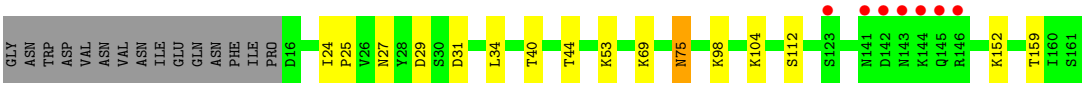
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	254	254	254	0	0
2	B	212	212	212	0	0

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.

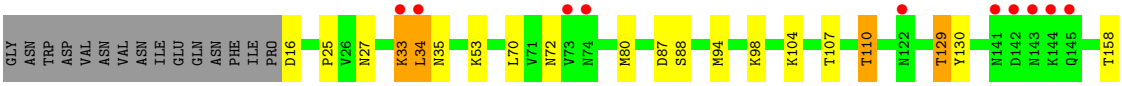
• Molecule 1: CS6 FIMBRIAL SUBUNIT B, CS6 FIMBRIAL SUBUNIT A

Chain A:



• Molecule 1: CS6 FIMBRIAL SUBUNIT B, CS6 FIMBRIAL SUBUNIT A

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	51.96Å 69.22Å 72.36Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	27.29 – 1.04 27.29 – 1.05	Depositor EDS
% Data completeness (in resolution range)	98.9 (27.29-1.04) 98.0 (27.29-1.05)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.23 (at 1.05Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, R_{free}	0.155 , 0.167 0.166 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	8.1	Xtriage
Anisotropy	0.385	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.48 , 50.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.014 for -h,l,k	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	4844	wwPDB-VP
Average B, all atoms (Å^2)	15.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 11.11% 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, 0.333 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	0.76	2/1135 (0.2%)	0.78	0/1540
1	B	0.80	2/1135 (0.2%)	0.83	3/1540 (0.2%)
All	All	0.78	4/2270 (0.2%)	0.80	3/3080 (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	2	0
1	B	3	0
All	All	5	0

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	112	SER	CB-OG	-7.25	1.32	1.42
1	B	110	THR	CB-CG2	-6.67	1.30	1.52
1	B	129	THR	CB-CG2	-6.47	1.30	1.52
1	A	152	LYS	CE-NZ	-6.06	1.33	1.49

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	34	LEU	CB-CG-CD1	-6.20	100.46	111.00
1	B	129	THR	CA-CB-CG2	5.99	120.79	112.40
1	B	129	THR	OG1-CB-CG2	5.67	123.05	110.00

All (5) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	A	40	THR	CB
1	A	159	THR	CB
1	B	107	THR	CB
1	B	110	THR	CB
1	B	129	THR	CB

There are no planarity outliers.

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	1113	1076	1075	21	2
1	B	1113	1076	1075	17	1
2	A	254	0	0	16	3
2	B	212	0	0	7	5
All	All	2692	2152	2150	38	6

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 9.

The worst 5 of 38 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:A:75:ASN:OD1	2:A:2168:HOH:O	1.84	0.95
1:A:44:THR:HB	2:A:2012:HOH:O	1.67	0.95
1:A:75:ASN:HB2	2:A:2160:HOH:O	1.68	0.90
1:A:75:ASN:HB2	2:A:2164:HOH:O	1.74	0.87
1:A:25:PRO:HG3	2:A:2252:HOH:O	1.76	0.85

The worst 5 of 6 symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:94:MET:CE	2:B:2211:HOH:O[3_544]	1.82	0.38
1:A:104:LYS:HD2	2:B:2199:HOH:O[4_545]	1.25	0.35

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:A:2193:HOH:O	2:A:2219:HOH:O[3_555]	1.85	0.35
2:A:2103:HOH:O	2:B:2053:HOH:O[3_554]	1.87	0.33
2:A:2243:HOH:O	2:B:2143:HOH:O[4_445]	1.90	0.30

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	144/161 (89%)	143 (99%)	1 (1%)	0	100	100
1	B	144/161 (89%)	141 (98%)	3 (2%)	0	100	100
All	All	288/322 (89%)	284 (99%)	4 (1%)	0	100	100

There are no Ramachandran outliers to report.

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	126/140 (90%)	124 (98%)	2 (2%)	62	27
1	B	126/140 (90%)	124 (98%)	2 (2%)	62	27
All	All	252/280 (90%)	248 (98%)	4 (2%)	62	27

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	34	LEU
1	A	75	ASN
1	B	33	LYS
1	B	129	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	75	ASN
1	A	143	ASN
1	B	37	GLN
1	B	65	GLN
1	B	74	ASN

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	146/161 (90%)	0.32	7 (4%)	30	27	5, 10, 23, 30	0
1	B	146/161 (90%)	0.40	10 (6%)	17	18	6, 11, 25, 32	0
All	All	292/322 (90%)	0.36	17 (5%)	23	22	5, 11, 24, 32	0

The worst 5 of 17 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	141	ASN	6.0
1	B	143	ASN	5.7
1	B	144	LYS	5.5
1	B	142	ASP	5.3
1	B	145	GLN	4.4

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.