wwPDB X-ray Structure Validation Summary Report (i)

May 25, 2020 – 05:40 am BST

PDB ID	:	5O7H
Title	:	Structure of the Cascade-I-Fv complex from Shewanella putrefaciens
Authors	:	Pausch, P.; Altegoer, F.; Bange, G.
Deposited on	:	2017-06-08
Resolution	:	3.00 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 following versions of software and data (see references (1)) were used in the production of this report:

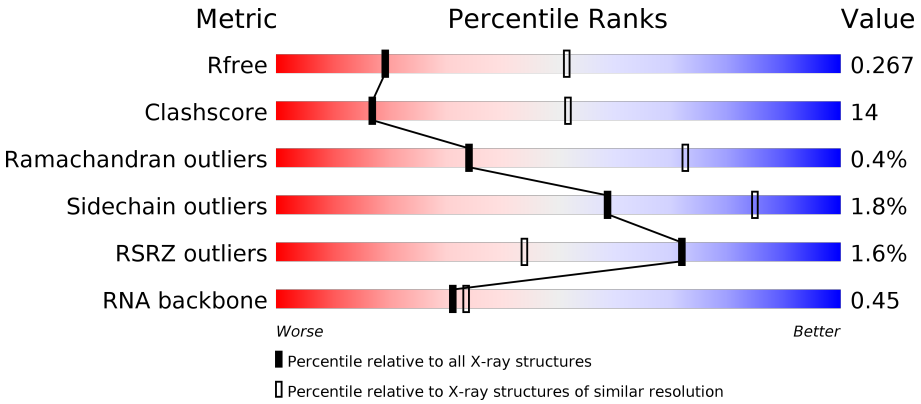
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 3.00 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.



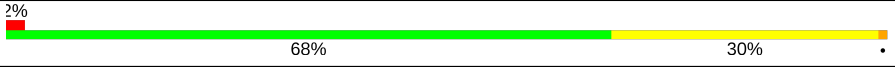
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)
RNA backbone	3102	1173 (3.30-2.70)

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 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	43	2% 53% 35% 12%
2	B	167	4% 50% 26% • 23%
3	C	315	66% 22% • 12%
3	D	315	% 71% 28% •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	E	315	
4	F	336	

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 12262 atoms, of which 456 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 RNA chain called crRNA.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	43	Total	C	H	N	O	P	0	0	0
			1353	400	456	164	290	43			

- Molecule 2 is a protein called CRISPR-associated protein, Csy4 family.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	129	Total	C	N	O	Se	0	0	0
			1042	669	183	187	3			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	15	MSE	-	initiating methionine	UNP A4Y6G3

- Molecule 3 is a protein called Cas7fv.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	C	278	Total	C	N	O	S	Se	0	0	0
			2218	1406	381	424	3	4			
3	D	315	Total	C	N	O	S	Se	0	0	0
			2506	1584	429	485	3	5			
3	E	312	Total	C	N	O	S	Se	0	0	0
			2484	1570	425	482	3	4			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	1	MSE	-	initiating methionine	UNP A4Y6G1
D	1	MSE	-	initiating methionine	UNP A4Y6G1
E	1	MSE	-	initiating methionine	UNP A4Y6G1

- Molecule 4 is a protein called Cas5fv.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
4	F	336	Total	C	N	O	S	Se	0	0	0
			2659	1695	440	510	4	10			

There is a discrepancy between the modelled and reference sequences:

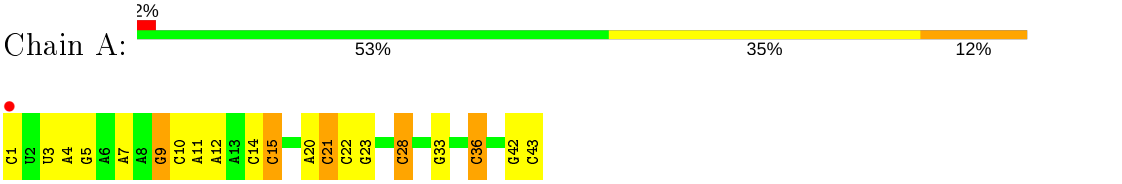
Chain	Residue	Modelled	Actual	Comment	Reference
F	1	MSE	-	initiating methionine	UNP A4Y6G2

3 Residue-property plots

These plots are drawn for all protein, RNA and DNA 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.

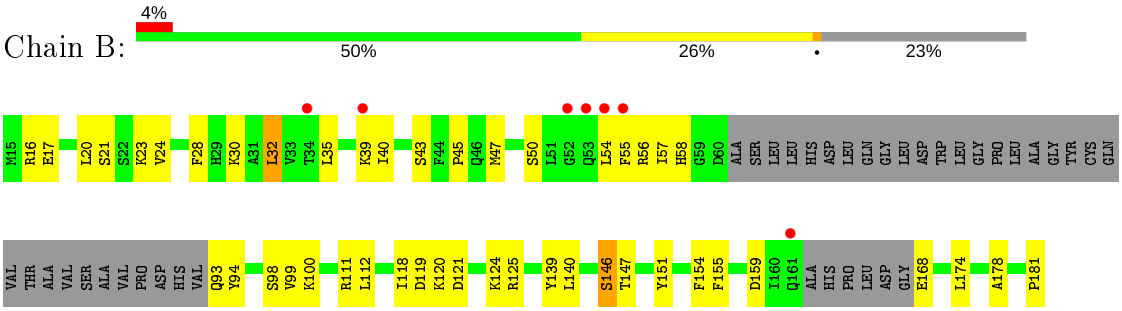
- Molecule 1: crRNA

Chain A:



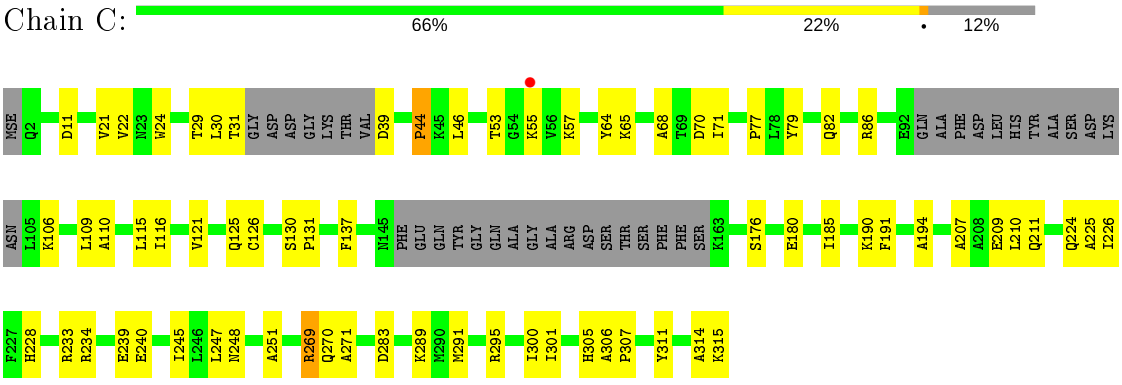
- Molecule 2: CRISPR-associated protein, Csy4 family

Chain B:



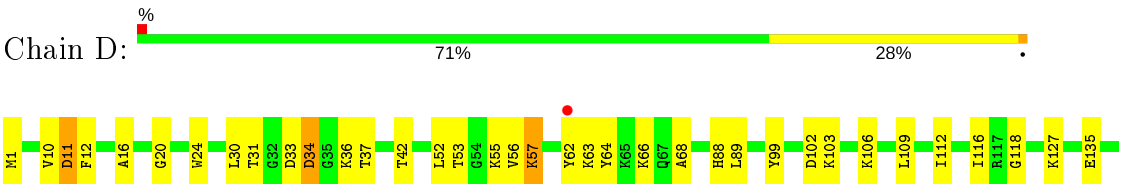
- Molecule 3: Cas7fv

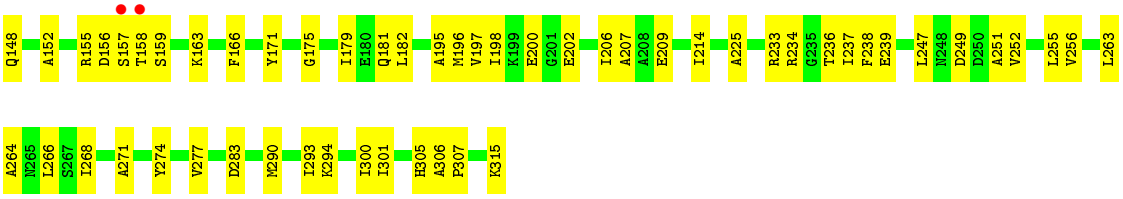
Chain C:



- Molecule 3: Cas7fv

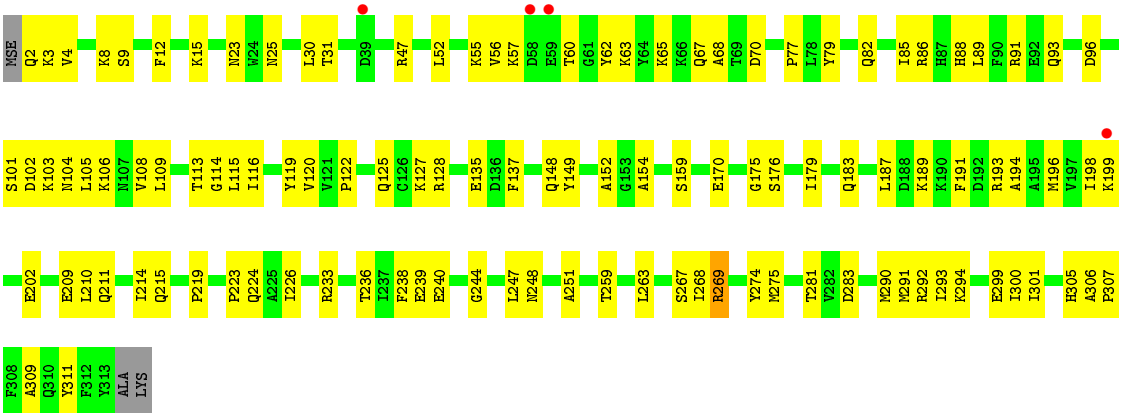
Chain D:





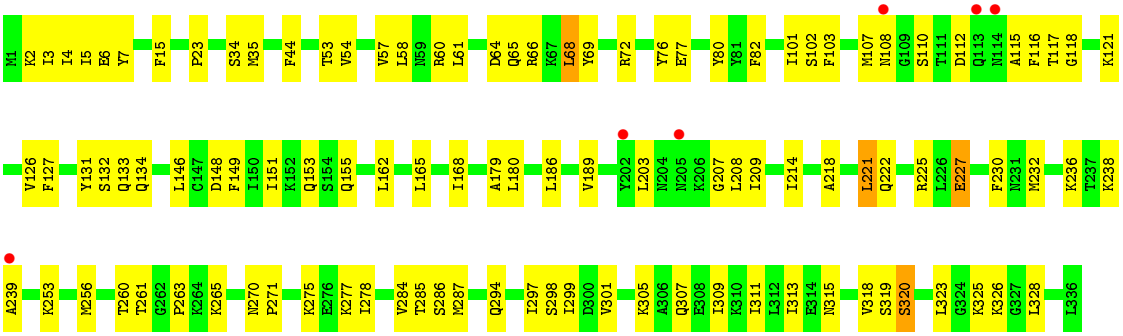
- Molecule 3: Cas7fv

Chain E:



- Molecule 4: Cas5fv

Chain F:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	157.00Å 65.89Å 160.68Å 90.00° 98.61° 90.00°	Depositor
Resolution (Å)	49.29 – 3.00 49.29 – 3.00	Depositor EDS
% Data completeness (in resolution range)	99.8 (49.29-3.00) 99.8 (49.29-3.00)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.96 (at 3.01Å)	Xtriage
Refinement program	PHENIX (1.11.1_2575: ???)	Depositor
R, R_{free}	0.221 , 0.267 0.221 , 0.267	Depositor DCC
R_{free} test set	1617 reflections (4.92%)	wwPDB-VP
Wilson B-factor (Å^2)	78.4	Xtriage
Anisotropy	0.292	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 57.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.019 for l,-k,h	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	12262	wwPDB-VP
Average B, all atoms (Å^2)	77.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.74% 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

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	0.21	0/1002	0.72	0/1559
2	B	0.25	0/1058	0.41	0/1405
3	C	0.24	0/2252	0.43	0/3029
3	D	0.25	0/2550	0.43	0/3434
3	E	0.25	0/2528	0.42	0/3406
4	F	0.25	0/2694	0.42	0/3608
All	All	0.24	0/12084	0.46	0/16441

There are no bond length outliers.

There are no bond angle outliers.

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	897	456	456	42	0
2	B	1042	0	1069	37	0
3	C	2218	0	2211	48	0
3	D	2506	0	2468	73	0
3	E	2484	0	2438	99	0
4	F	2659	0	2698	86	0
All	All	11806	456	11340	323	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 14.

The worst 5 of 323 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)
3:C:31:THR:HG1	3:C:39:ASP:N	1.38	1.19
3:E:4:VAL:HG12	3:E:8:LYS:HZ1	1.38	0.89
3:E:283:ASP:HB2	3:E:300:ILE:HG23	1.58	0.86
3:E:56:VAL:HG12	3:E:57:LYS:H	1.42	0.85
4:F:275:LYS:HG2	4:F:285:THR:HG22	1.56	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	
2	B	123/167 (74%)	113 (92%)	9 (7%)	1 (1%)	19	57
3	C	270/315 (86%)	255 (94%)	14 (5%)	1 (0%)	34	72
3	D	313/315 (99%)	300 (96%)	12 (4%)	1 (0%)	41	76
3	E	310/315 (98%)	300 (97%)	9 (3%)	1 (0%)	41	76
4	F	334/336 (99%)	307 (92%)	26 (8%)	1 (0%)	41	76
All	All	1350/1448 (93%)	1275 (94%)	70 (5%)	5 (0%)	34	72

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	D	34	ASP
4	F	320	SER
2	B	45	PRO
3	E	194	ALA
3	C	44	PRO

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
2	B	115/141 (82%)	109 (95%)	6 (5%)	23 59
3	C	243/267 (91%)	241 (99%)	2 (1%)	81 93
3	D	272/267 (102%)	269 (99%)	3 (1%)	73 90
3	E	270/267 (101%)	267 (99%)	3 (1%)	73 90
4	F	296/286 (104%)	289 (98%)	7 (2%)	49 79
All	All	1196/1228 (97%)	1175 (98%)	21 (2%)	59 85

5 of 21 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	57	LYS
3	E	2	GLN
4	F	148	ASP
3	D	11	ASP
4	F	221	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
2	B	58	HIS
3	C	82	GLN
3	C	270	GLN
4	F	176	GLN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	A	41/43 (95%)	5 (12%)	0

All (5) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	A	9	G
1	A	15	C
1	A	21	C
1	A	28	C
1	A	36	C

There are no RNA pucker outliers to report.

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 carbohydrates 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	43/43 (100%)	0.22	1 (2%) 60 31	46, 85, 126, 151	0
2	B	126/167 (75%)	0.42	7 (5%) 24 8	75, 118, 141, 156	0
3	C	274/315 (86%)	-0.24	1 (0%) 92 79	50, 69, 115, 137	0
3	D	310/315 (98%)	-0.26	3 (0%) 82 59	40, 61, 121, 166	0
3	E	308/315 (97%)	-0.23	4 (1%) 77 51	40, 76, 110, 168	0
4	F	326/336 (97%)	-0.17	6 (1%) 68 40	49, 82, 120, 153	0
All	All	1387/1491 (93%)	-0.15	22 (1%) 72 44	40, 75, 127, 168	0

The worst 5 of 22 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	D	157	SER	5.7
3	E	59	GLU	5.4
3	E	58	ASP	4.9
2	B	55	PHE	4.1
1	A	1	C	3.7

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 (i)

There are no such residues in this entry.