wwPDB X-ray Structure Validation Summary Report (i)

Nov 14, 2023 – 11:39 PM JST

PDB ID	:	6IFN
Title	:	Crystal structure of Type III-A CRISPR Csm complex
Authors	:	You, L.; Wang, J.; Wang, Y.
Deposited on	:	2018-09-20
Resolution	:	2.90 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 (i)) were used in the production of this report:

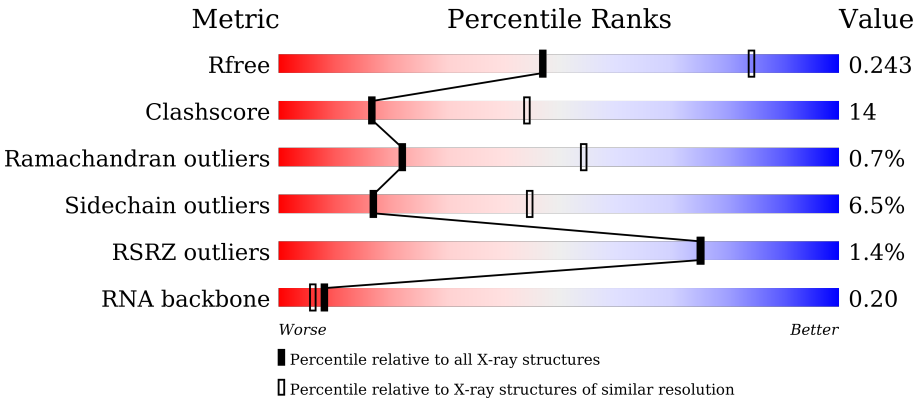
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 2.90 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.



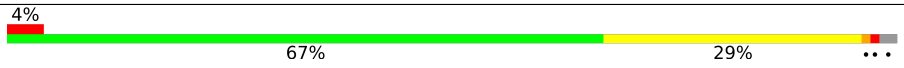
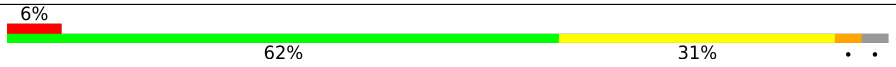
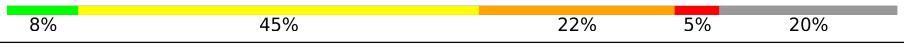
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)
RNA backbone	3102	1007 (3.16-2.64)

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	758	69% 25%
2	E	220	% 72% 25%
2	F	220	70% 29%
2	G	220	% 78% 21%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	B	299	
4	C	126	
4	D	126	
5	H	357	
6	N	40	

2 Entry composition (i)

There are 8 unique types of molecules in this entry. The entry contains 18048 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 Type III-A CRISPR-associated protein Csm1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	733	Total	C	N	O	S	0	0	0
			5726	3675	959	1076	16			

- Molecule 2 is a protein called Type III-A CRISPR-associated RAMP protein Csm3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	F	220	Total	C	N	O	S	0	0	0
			1681	1063	287	328	3			
2	E	219	Total	C	N	O	S	0	0	0
			1675	1059	286	327	3			
2	G	220	Total	C	N	O	S	0	0	0
			1621	1022	281	316	2			

- Molecule 3 is a protein called Type III-A CRISPR-associated RAMP protein Csm4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	B	297	Total	C	N	O	S	0	0	0
			2311	1482	383	441	5			

- Molecule 4 is a protein called Type III-A CRISPR-associated protein Csm2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	C	124	Total	C	N	O	S	0	0	0
			959	622	160	175	2			
4	D	123	Total	C	N	O	S	0	0	0
			904	588	149	165	2			

- Molecule 5 is a protein called Type III-A CRISPR-associated RAMP protein Csm5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	H	345	Total	C	N	O	S	0	0	0
			2499	1600	433	462	4			

- Molecule 6 is a RNA chain called RNA (32-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	N	32	Total	C	N	O	P	0	0	0
			670	302	115	222	31			

- Molecule 7 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	1	Total	Zn	0	0
			1	1		

- Molecule 8 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

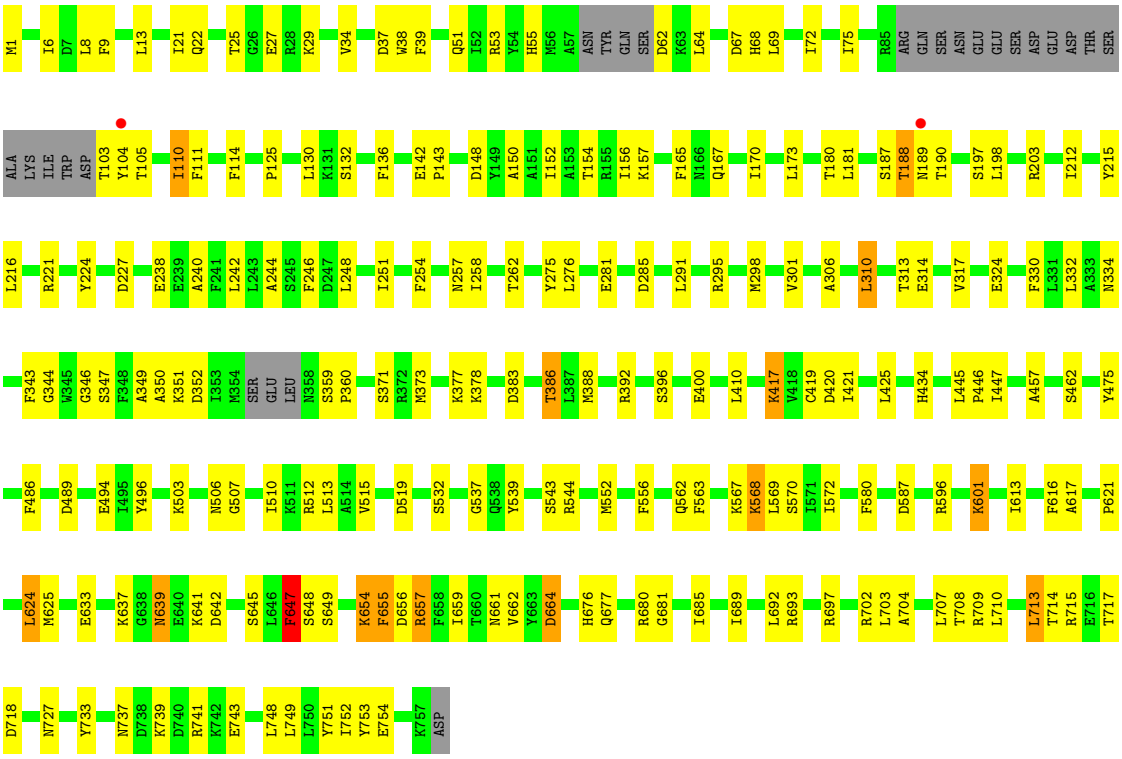
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	1	Total	Mn	0	0
			1	1		

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.

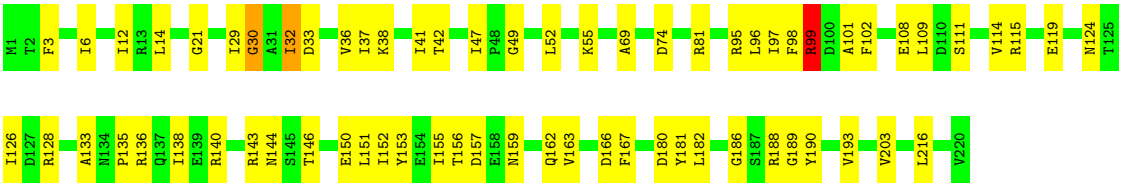
• Molecule 1: Type III-A CRISPR-associated protein Csm1

Chain A:



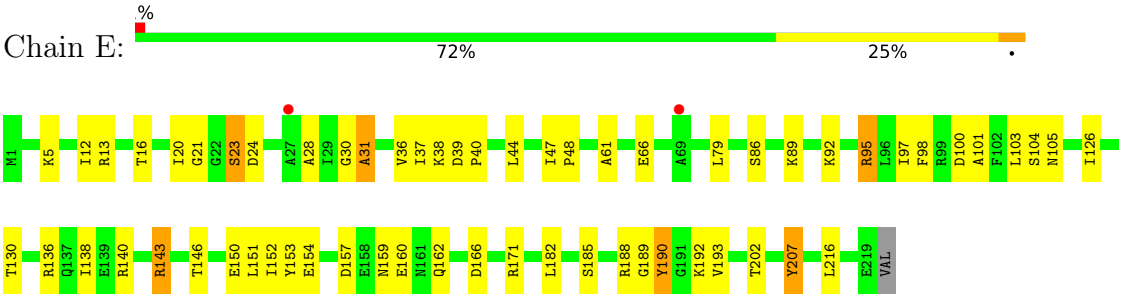
• Molecule 2: Type III-A CRISPR-associated RAMP protein Csm3

Chain F:



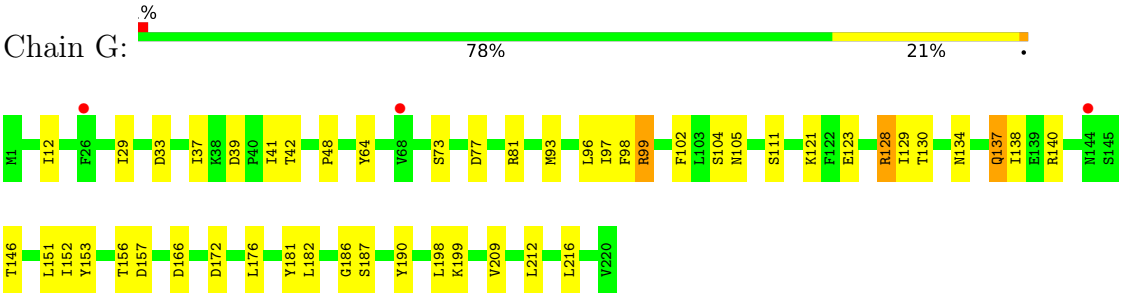
• Molecule 2: Type III-A CRISPR-associated RAMP protein Csm3

Chain E:



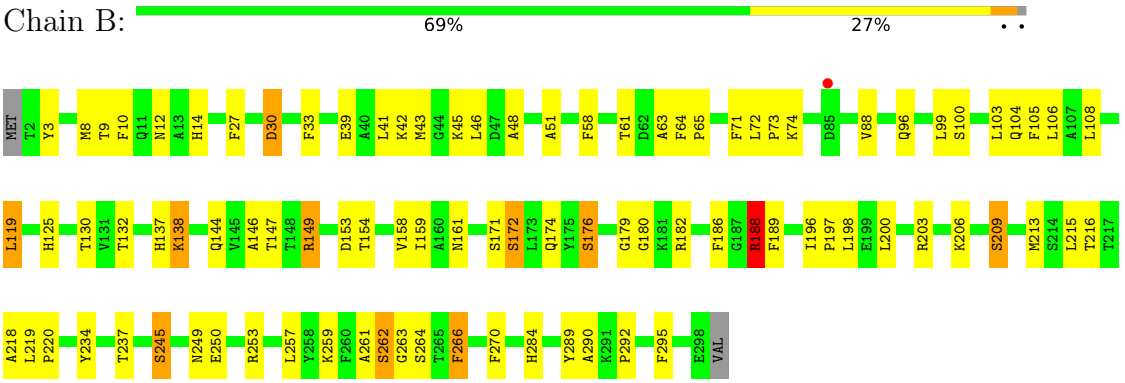
• Molecule 2: Type III-A CRISPR-associated RAMP protein Csm3

Chain G:



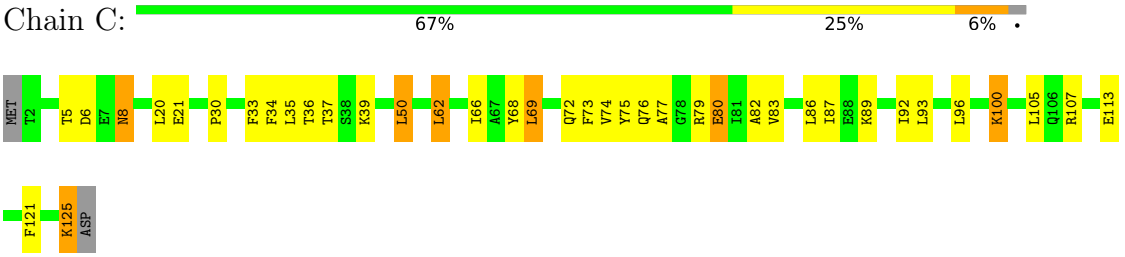
• Molecule 3: Type III-A CRISPR-associated RAMP protein Csm4

Chain B:



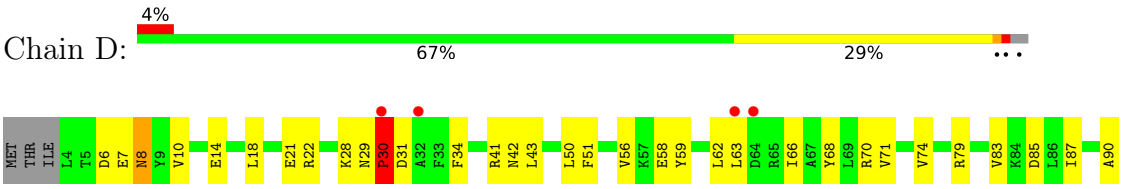
• Molecule 4: Type III-A CRISPR-associated protein Csm2

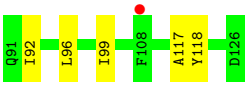
Chain C:



• Molecule 4: Type III-A CRISPR-associated protein Csm2

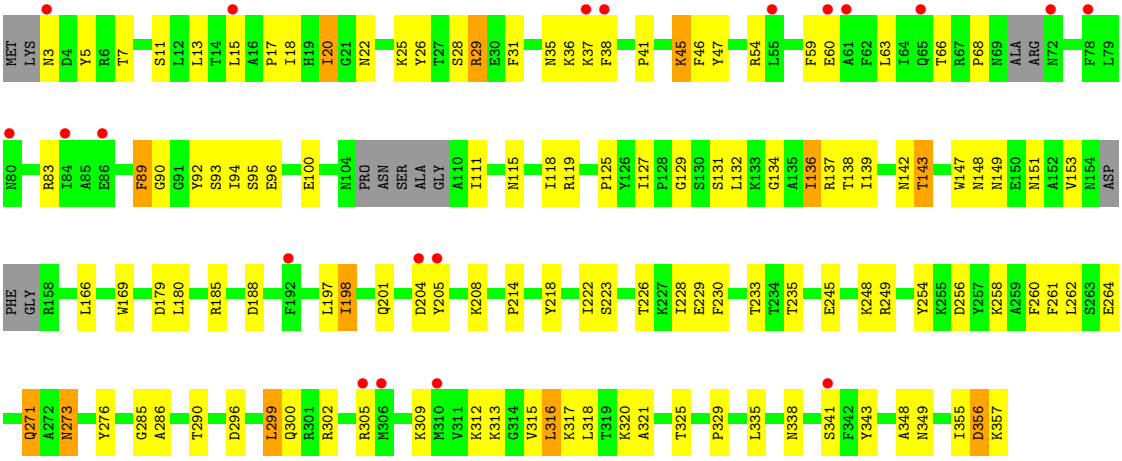
Chain D:





• Molecule 5: Type III-A CRISPR-associated RAMP protein Csm5

Chain H:



• Molecule 6: RNA (32-MER)

Chain N:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	123.54Å 82.33Å 161.36Å 90.00° 99.32° 90.00°	Depositor
Resolution (Å)	48.68 – 2.90 48.64 – 2.88	Depositor EDS
% Data completeness (in resolution range)	83.8 (48.68-2.90) 83.8 (48.64-2.88)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.45 (at 2.86Å)	Xtriage
Refinement program	REFMAC 5.8.0230	Depositor
R, R_{free}	0.201 , 0.244 0.203 , 0.243	Depositor DCC
R_{free} test set	3053 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	50.6	Xtriage
Anisotropy	0.090	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 40.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.44$, $< L^2 > = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	18048	wwPDB-VP
Average B, all atoms (Å^2)	53.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.63% 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

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, MN

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	0.58	0/5846	0.73	2/7909 (0.0%)
2	E	0.55	0/1699	0.73	0/2295
2	F	0.52	0/1706	0.69	1/2304 (0.0%)
2	G	0.47	0/1644	0.67	0/2228
3	B	0.66	0/2363	0.76	1/3202 (0.0%)
4	C	0.53	0/973	0.69	2/1316 (0.2%)
4	D	0.50	0/918	0.72	1/1250 (0.1%)
5	H	0.55	0/2551	0.76	0/3464
6	N	0.68	1/747 (0.1%)	1.15	10/1160 (0.9%)
All	All	0.57	1/18447 (0.0%)	0.75	17/25128 (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	0	1
2	E	0	1
2	F	0	2
2	G	0	1
3	B	0	3
4	C	0	1
4	D	0	1
5	H	0	2
All	All	0	12

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	N	10	C	O3'-P	-5.23	1.54	1.61

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	648	SER	N-CA-CB	9.01	124.01	110.50
6	N	21	G	O4'-C1'-N9	7.04	113.83	108.20
6	N	10	C	O5'-P-OP1	-6.91	99.48	105.70
6	N	4	G	O5'-P-OP2	-6.90	99.49	105.70
4	C	8	ASN	N-CA-C	6.77	129.28	111.00

There are no chirality outliers.

5 of 12 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	680	ARG	Sidechain
3	B	179	GLY	Peptide
3	B	182	ARG	Sidechain
2	F	128	ARG	Sidechain
2	F	99	ARG	Sidechain

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	5726	0	5388	139	0
2	E	1675	0	1642	57	0
2	F	1681	0	1632	61	0
2	G	1621	0	1528	41	0
3	B	2311	0	2213	61	1
4	C	959	0	926	39	0
4	D	904	0	819	32	0
5	H	2499	0	2227	106	0
6	N	670	0	345	36	0
7	A	1	0	0	0	0
8	A	1	0	0	0	0
All	All	18048	0	16720	494	1

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 494 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)
5:H:198:ILE:HD11	5:H:223:SER:CB	1.23	1.57
5:H:198:ILE:CD1	5:H:223:SER:CB	2.17	1.23
5:H:38:PHE:N	5:H:94:ILE:O	1.76	1.17
1:A:1:MET:N	1:A:167:GLN:OE1	1.77	1.17
5:H:37:LYS:HA	5:H:95:SER:HA	1.28	1.14

All (1) 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)
3:B:209:SER:OG	3:B:250:GLU:OE1[2_557]	2.07	0.13

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	725/758 (96%)	688 (95%)	34 (5%)	3 (0%)	34 66
2	E	217/220 (99%)	206 (95%)	9 (4%)	2 (1%)	17 48
2	F	218/220 (99%)	208 (95%)	9 (4%)	1 (0%)	29 61
2	G	218/220 (99%)	210 (96%)	8 (4%)	0	100 100
3	B	295/299 (99%)	282 (96%)	13 (4%)	0	100 100
4	C	122/126 (97%)	118 (97%)	3 (2%)	1 (1%)	19 51
4	D	121/126 (96%)	112 (93%)	8 (7%)	1 (1%)	19 51
5	H	337/357 (94%)	291 (86%)	39 (12%)	7 (2%)	7 26
All	All	2253/2326 (97%)	2115 (94%)	123 (6%)	15 (1%)	22 54

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	F	30	GLY
4	D	30	PRO
5	H	68	PRO
5	H	329	PRO
2	E	207	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	564/651 (87%)	527 (93%)	37 (7%)	16	44
2	E	173/188 (92%)	164 (95%)	9 (5%)	23	55
2	F	172/188 (92%)	168 (98%)	4 (2%)	50	80
2	G	156/188 (83%)	147 (94%)	9 (6%)	20	50
3	B	243/263 (92%)	228 (94%)	15 (6%)	18	47
4	C	91/112 (81%)	85 (93%)	6 (7%)	16	44
4	D	77/112 (69%)	69 (90%)	8 (10%)	7	21
5	H	217/312 (70%)	195 (90%)	22 (10%)	7	23
All	All	1693/2014 (84%)	1583 (94%)	110 (6%)	17	45

5 of 110 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	E	23	SER
4	D	6	ASP
5	H	341	SER
5	H	197	LEU
2	E	66	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
2	E	162	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
4	C	72	GLN
5	H	201	GLN
5	H	148	ASN
3	B	144	GLN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
6	N	31/40 (77%)	12 (38%)	3 (9%)

5 of 12 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
6	N	2	C
6	N	3	G
6	N	9	G
6	N	10	C
6	N	15	U

All (3) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
6	N	7	A
6	N	13	U
6	N	25	U

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)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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	733/758 (96%)	-0.56	2 (0%) 94 94	22, 43, 79, 113	0
2	E	219/220 (99%)	-0.41	2 (0%) 84 84	26, 43, 85, 135	0
2	F	220/220 (100%)	-0.28	0 100 100	34, 55, 84, 129	0
2	G	220/220 (100%)	-0.03	3 (1%) 75 75	41, 65, 96, 122	0
3	B	297/299 (99%)	-0.59	1 (0%) 94 94	18, 32, 61, 119	0
4	C	124/126 (98%)	-0.44	0 100 100	35, 55, 80, 101	0
4	D	123/126 (97%)	0.08	5 (4%) 37 32	49, 72, 101, 127	0
5	H	345/357 (96%)	0.10	20 (5%) 23 19	40, 70, 106, 129	0
6	N	32/40 (80%)	-0.35	0 100 100	20, 43, 79, 118	0
All	All	2313/2366 (97%)	-0.33	33 (1%) 75 75	18, 52, 93, 135	0

The worst 5 of 33 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	D	30	PRO	5.4
5	H	3	ASN	5.1
5	H	310	MET	4.6
3	B	85	ASP	4.3
5	H	78	PHE	4.1

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 monosaccharides in this entry.

6.4 Ligands

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
8	MN	A	802	1/1	0.99	0.12	66,66,66,66	0
7	ZN	A	801	1/1	1.00	0.09	36,36,36,36	0

6.5 Other polymers

There are no such residues in this entry.