wwPDB X-ray Structure Validation Summary Report (i)

Feb 19, 2016 – 09:46 PM GMT

PDB ID	:	5AWG
Title	:	Crystal structure of Hg-bound SufB-SufC-SufD complex from Escherichia coli
Authors	:	Hirabayashi, K.; Wada, K.
Deposited on	:	2015-07-03
Resolution	:	4.28 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:

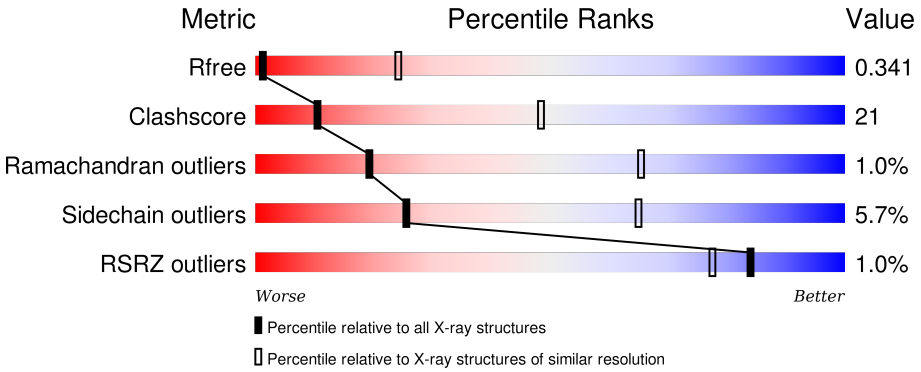
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026982
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20026982

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 4.28 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	1043 (4.92-3.60)
Clashscore	102246	1145 (4.92-3.60)
Ramachandran outliers	100387	1088 (4.92-3.60)
Sidechain outliers	100360	1072 (4.92-3.60)
RSRZ outliers	91569	1047 (4.92-3.60)

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	495	% 41% 32% • 23%
1	E	495	% 33% 37% 5% 25%
2	B	423	% 56% 27% • 15%
2	F	423	% 45% 35% • • 16%
3	C	248	62% 29% • • •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	D	248	62% 30% 5%
3	G	248	2% 61% 32%
3	H	248	% 58% 33% 5%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 18808 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 FeS cluster assembly protein SufB.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	380	Total	C	N	O	S	0	0	0
			2968	1869	509	572	18			
1	E	372	Total	C	N	O	S	0	0	0
			2910	1833	498	561	18			

- Molecule 2 is a protein called FeS cluster assembly protein SufD.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	361	Total	C	N	O	S	0	0	0
			2798	1743	514	533	8			
2	F	355	Total	C	N	O	S	0	0	0
			2752	1712	507	525	8			

- Molecule 3 is a protein called Probable ATP-dependent transporter SufC.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	237	Total	C	N	O	S	0	0	0
			1846	1169	312	358	7			
3	D	235	Total	C	N	O	S	0	0	0
			1829	1158	310	354	7			
3	G	239	Total	C	N	O	S	0	0	0
			1864	1179	315	363	7			
3	H	236	Total	C	N	O	S	0	0	0
			1837	1164	311	355	7			

- Molecule 4 is MERCURY (II) ION (three-letter code: HG) (formula: Hg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Hg	0	0
			1	1		
4	A	1	Total	Hg	0	0
			1	1		

Continued on next page...

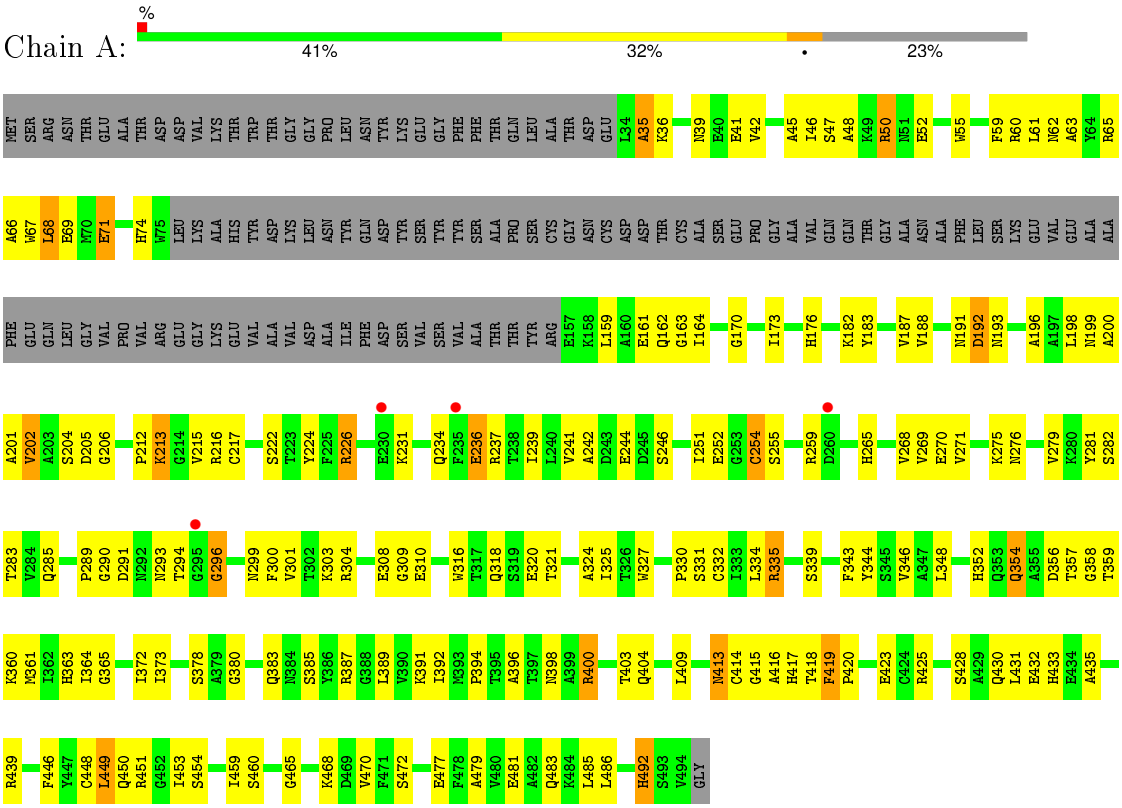
Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	F	1	Total Hg 1 1	0	0
4	E	1	Total Hg 1 1	0	0

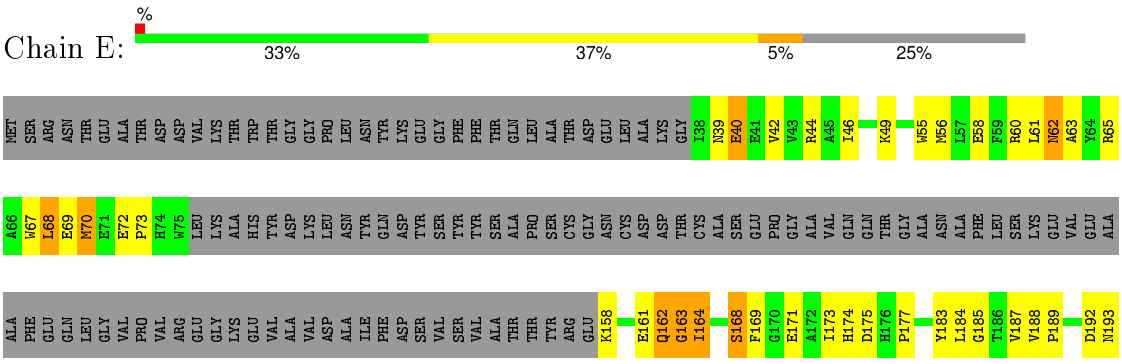
3 Residue-property plots (i)

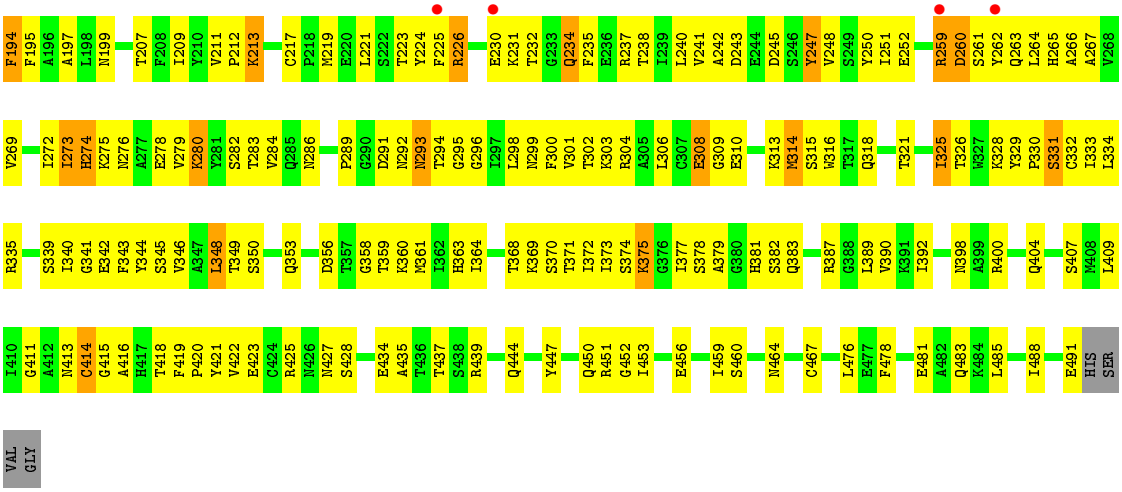
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according 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: FeS cluster assembly protein SufB



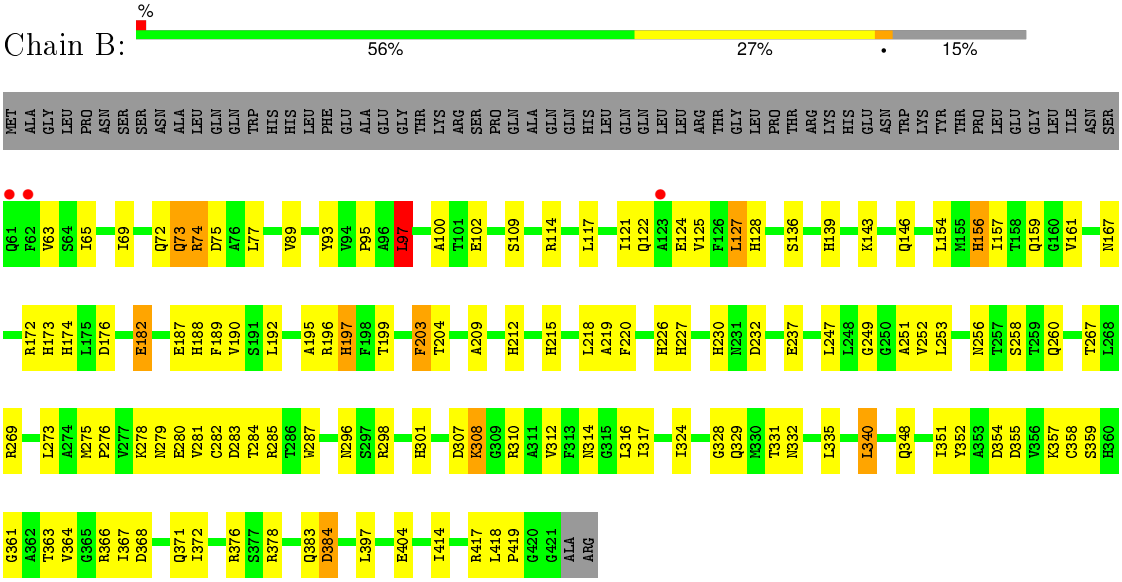
- Molecule 1: FeS cluster assembly protein SufB





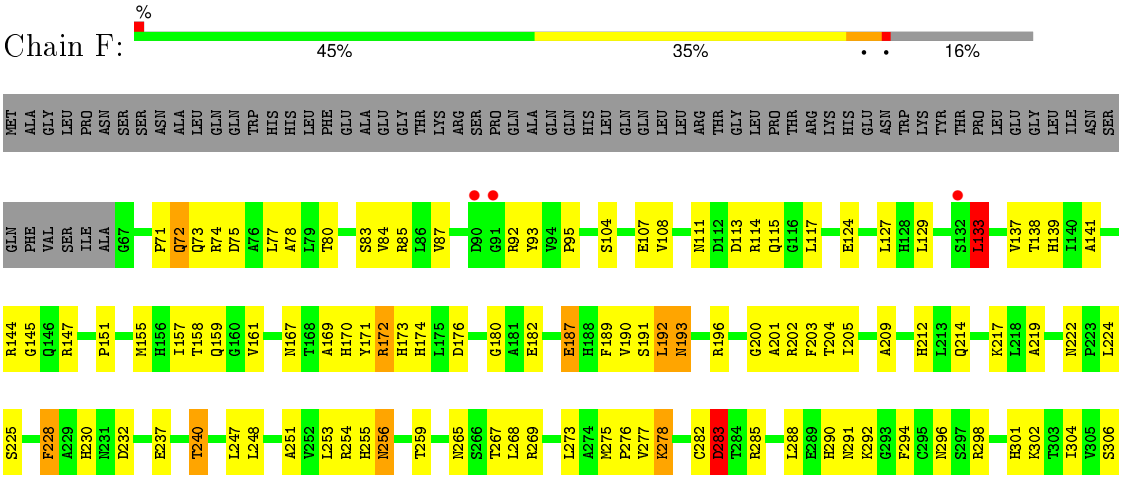
• Molecule 2: FeS cluster assembly protein SufD

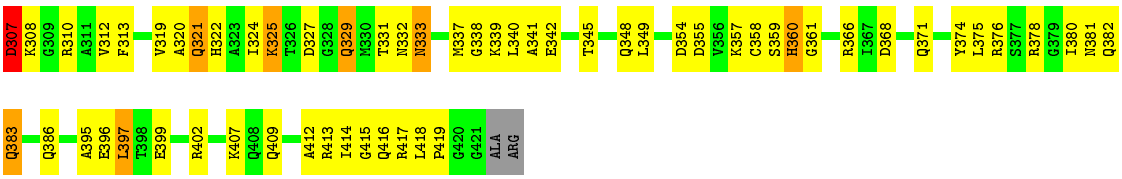
Chain B:



• Molecule 2: FeS cluster assembly protein SufD

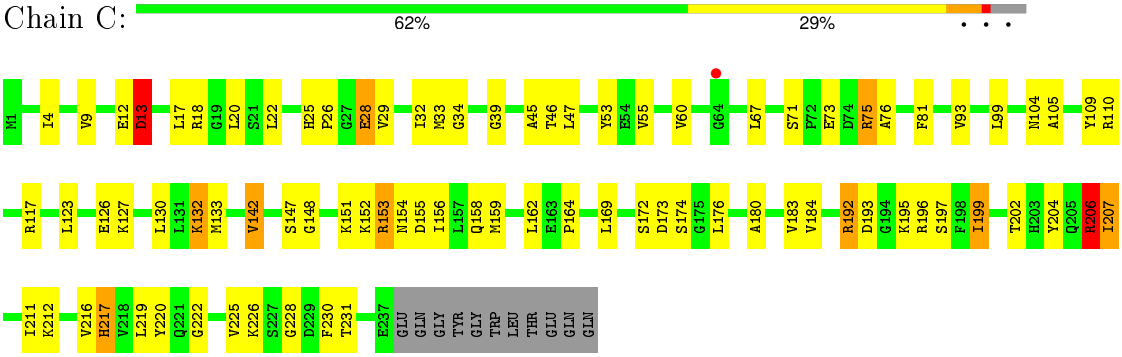
Chain F:





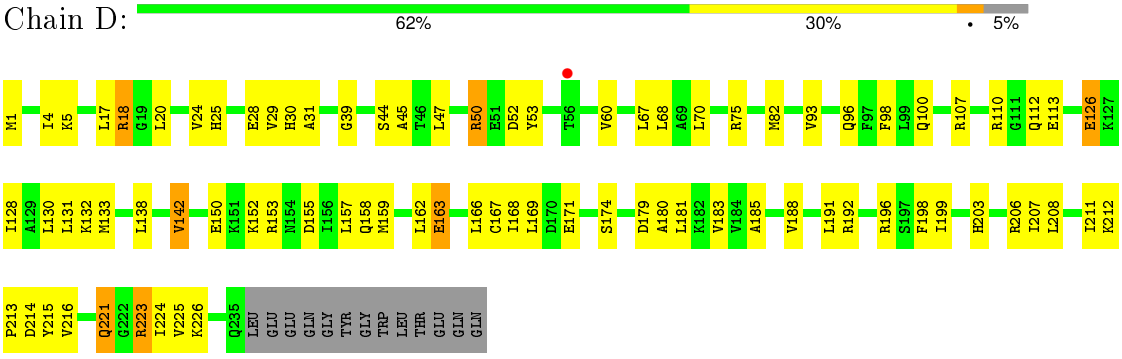
- Molecule 3: Probable ATP-dependent transporter SufC

Chain C:



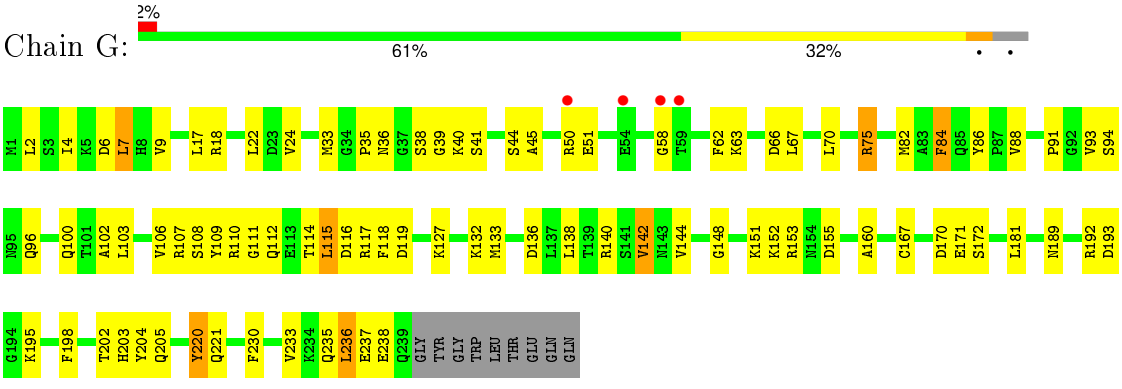
- Molecule 3: Probable ATP-dependent transporter SufC

Chain D:



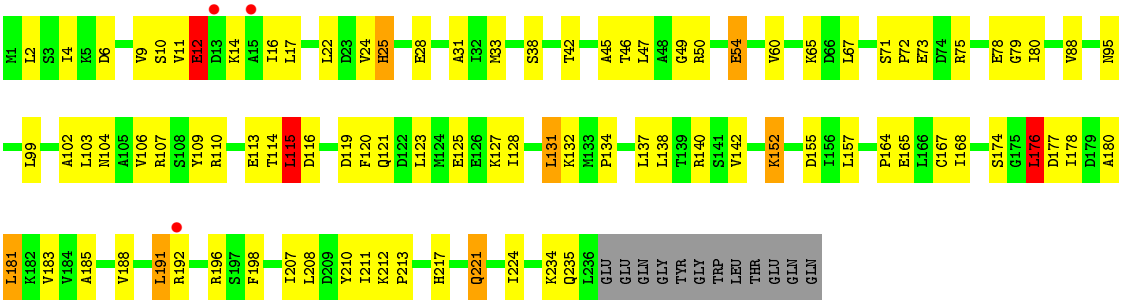
- Molecule 3: Probable ATP-dependent transporter SufC

Chain G:



- Molecule 3: Probable ATP-dependent transporter SufC

Chain H:


M1 L2 S3 I4 K5 D6 V9 S10 V11 E12 D13 K14 A15 I16 L17 L22 D23 V24 H25 E28 A31 I32 M33 S38 T42 A45 T46 L47 A48 G49 R50 E54 V60 K65 D66 L67 S71 P72 E73 D74 R75 E78 G79 I80 V88 N95
L99 A102 L103 N104 A105 V106 R107 S108 Y109 R110 E113 T114 L115 D116 D119 F120 Q121 D122 L123 M124 E125 E126 K127 I128 L131 K132 M133 P134 L137 L138 T139 R140 S141 V142 K152 D155 I156 L157 P164 E165 L166 C167 I168 S174 G175 L176 D177 I178 D179 A180
L181 K182 V183 V184 A185 V188 L191 R192 R196 S197 F198 I207 L208 D209 Y210 I211 K212 P213 H217 Q221 I224 K234 Q235 L236 GLU GLU GLN GLY TYR GLY TRP LEU THR GLU GLN GLN

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	119.85Å 139.38Å 124.41Å 90.00° 113.55° 90.00°	Depositor
Resolution (Å)	43.94 – 4.28 47.93 – 4.28	Depositor EDS
% Data completeness (in resolution range)	98.4 (43.94-4.28) 92.9 (47.93-4.28)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.50 (at 4.29Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, R_{free}	0.295 , 0.340 0.295 , 0.341	Depositor DCC
R_{free} test set	2510 reflections (5.01%)	DCC
Wilson B-factor (Å^2)	95.8	Xtriage
Anisotropy	0.399	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.26 , 34.5	EDS
Estimated twinning fraction	0.032 for l,-k,h	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.29$	Xtriage
Outliers	1 of 25606 reflections (0.004%)	Xtriage
F_o,F_c correlation	0.81	EDS
Total number of atoms	18808	wwPDB-VP
Average B, all atoms (Å^2)	110.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.46% 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

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

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	0.37	0/3031	0.76	4/4104 (0.1%)
1	E	0.40	0/2972	0.84	2/4025 (0.0%)
2	B	0.31	0/2847	0.69	2/3858 (0.1%)
2	F	0.34	0/2800	0.82	5/3794 (0.1%)
3	C	0.34	0/1875	0.84	3/2529 (0.1%)
3	D	0.31	0/1858	0.76	0/2506
3	G	0.32	0/1893	0.69	2/2553 (0.1%)
3	H	0.33	0/1866	0.90	10/2517 (0.4%)
All	All	0.34	0/19142	0.79	28/25886 (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	E	0	2
2	F	0	1
3	D	0	1
3	G	0	2
3	H	0	1
All	All	0	7

There are no bond length outliers.

The worst 5 of 28 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	68	LEU	CA-CB-CG	7.68	132.96	115.30
1	A	68	LEU	CB-CG-CD1	-7.41	98.41	111.00
2	B	97	LEU	CA-CB-CG	7.18	131.83	115.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	H	115	LEU	CA-CB-CG	7.01	131.42	115.30
2	F	192	LEU	CB-CA-C	-6.99	96.93	110.20

There are no chirality outliers.

5 of 7 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	D	113	GLU	Peptide
1	E	231	LYS	Peptide
1	E	61	LEU	Peptide
2	F	321	GLN	Sidechain
3	G	115	LEU	Peptide

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	2968	0	2876	150	0
1	E	2910	0	2816	184	0
2	B	2798	0	2754	89	0
2	F	2752	0	2707	150	0
3	C	1846	0	1857	73	0
3	D	1829	0	1840	57	0
3	G	1864	0	1871	59	0
3	H	1837	0	1851	80	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
4	E	1	0	0	0	0
4	F	1	0	0	0	0
All	All	18808	0	18572	798	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 21.

The worst 5 of 798 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:E:161:GLU:HA	1:E:162:GLN:HG3	1.46	0.98
2:F:230:HIS:HA	2:F:256:ASN:ND2	1.81	0.95
1:E:314:MET:SD	1:E:314:MET:N	2.43	0.91
2:B:249:GLY:HA2	2:B:417:ARG:HH22	1.36	0.91
1:A:316:TRP:HB3	1:A:318:GLN:HE22	1.34	0.90

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	376/495 (76%)	350 (93%)	18 (5%)	8 (2%)	9 52
1	E	368/495 (74%)	339 (92%)	26 (7%)	3 (1%)	24 70
2	B	359/423 (85%)	344 (96%)	14 (4%)	1 (0%)	46 83
2	F	353/423 (84%)	337 (96%)	14 (4%)	2 (1%)	30 74
3	C	235/248 (95%)	221 (94%)	11 (5%)	3 (1%)	15 60
3	D	233/248 (94%)	222 (95%)	10 (4%)	1 (0%)	39 80
3	G	237/248 (96%)	222 (94%)	12 (5%)	3 (1%)	15 60
3	H	234/248 (94%)	222 (95%)	10 (4%)	2 (1%)	21 67
All	All	2395/2828 (85%)	2257 (94%)	115 (5%)	23 (1%)	19 65

5 of 23 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	74	HIS
1	A	163	GLY
1	A	202	VAL
1	A	259	ARG
1	A	289	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	
1	A	320/413 (78%)	304 (95%)	16 (5%)	30	68
1	E	314/413 (76%)	283 (90%)	31 (10%)	10	42
2	B	297/350 (85%)	282 (95%)	15 (5%)	29	68
2	F	292/350 (83%)	270 (92%)	22 (8%)	17	56
3	C	203/212 (96%)	194 (96%)	9 (4%)	35	71
3	D	201/212 (95%)	192 (96%)	9 (4%)	34	71
3	G	205/212 (97%)	198 (97%)	7 (3%)	44	77
3	H	202/212 (95%)	195 (96%)	7 (4%)	43	76
All	All	2034/2374 (86%)	1918 (94%)	116 (6%)	25	65

5 of 116 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	162	GLN
1	E	260	ASP
3	G	235	GLN
1	E	168	SER
1	E	219	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 26 such sidechains are listed below:

Mol	Chain	Res	Type
3	C	158	GLN
1	E	293	ASN
2	F	360	HIS
3	D	154	ASN
1	E	193	ASN

5.3.3 RNA (i)

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

Of 4 ligands modelled in this entry, 4 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	380/495 (76%)	0.01	4 (1%)	82	75	62, 101, 137, 154	0
1	E	372/495 (75%)	0.04	4 (1%)	82	75	77, 111, 147, 178	0
2	B	361/423 (85%)	-0.14	3 (0%)	87	82	57, 96, 127, 142	0
2	F	355/423 (83%)	-0.14	3 (0%)	87	82	73, 104, 132, 149	0
3	C	237/248 (95%)	-0.13	1 (0%)	93	90	81, 110, 135, 156	0
3	D	235/248 (94%)	0.01	1 (0%)	93	90	74, 118, 149, 165	0
3	G	239/248 (96%)	-0.12	4 (1%)	73	63	77, 108, 142, 162	0
3	H	236/248 (95%)	0.07	3 (1%)	79	71	85, 137, 166, 171	0
All	All	2415/2828 (85%)	-0.05	23 (0%)	84	77	57, 108, 147, 178	0

The worst 5 of 23 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	91	GLY	3.5
2	B	61	GLN	3.3
2	F	132	SER	3.2
1	E	262	TYR	3.2
1	A	295	GLY	3.1

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 carbohydrates 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å^2)	Q<0.9
4	HG	F	600	1/1	0.94	0.10	-2.36	176,176,176,176	0
4	HG	E	600	1/1	0.97	0.06	-3.08	117,117,117,117	0
4	HG	A	600	1/1	0.97	0.14	-	128,128,128,128	0
4	HG	B	600	1/1	0.88	0.11	-	172,172,172,172	0

6.5 Other polymers

There are no such residues in this entry.