wwPDB X-ray Structure Validation Summary Report (i)

Aug 7, 2020 – 01:05 PM BST

PDB ID	:	6YVH
Title	:	CWC22-CWC27-EIF4A3 Complex
Authors	:	Basquin, J.; Busetto, V.; LeHir, H.; Conti, E.
Deposited on	:	2020-04-28
Resolution	:	3.19 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:

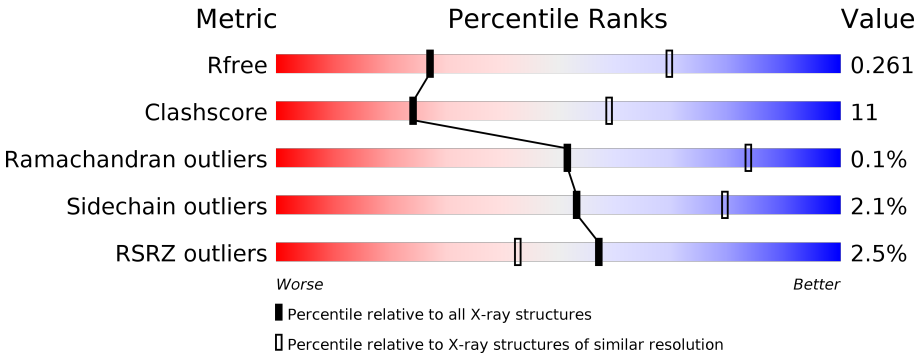
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.13.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.13.1

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 3.19 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.



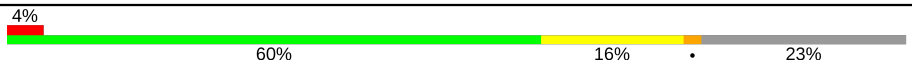
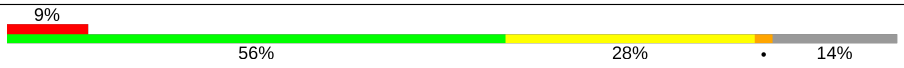
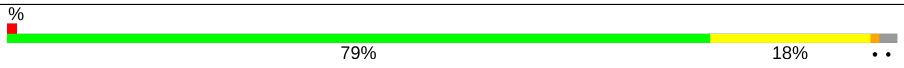
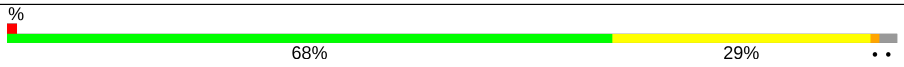
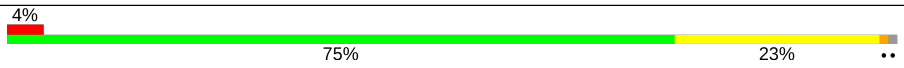
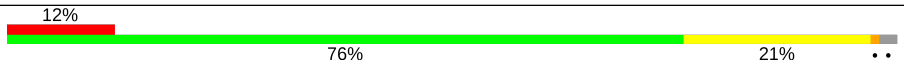
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1133 (3.20-3.20)
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)
RSRZ outliers	127900	1095 (3.20-3.20)

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	291	72% 22% • 6%
1	B	291	74% 20% 6%
1	D	291	% 70% 24% 6%
1	F	291	71% 21% • 7%
2	C	57	9% 61% 25% 14%
2	E	57	4% 70% 14% • 14%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	G	57	
2	I	57	
3	H	166	
3	J	166	
3	K	166	
3	L	166	

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 14536 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 Pre-mRNA-splicing factor CWC22 homolog.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	274	Total	C	N	O	S	0	1	0
			2074	1327	358	380	9			
1	B	274	Total	C	N	O	S	0	1	0
			2054	1318	348	380	8			
1	D	274	Total	C	N	O	S	0	1	0
			2068	1321	354	385	8			
1	F	272	Total	C	N	O	S	0	1	0
			2094	1338	353	393	10			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	116	ARG	-	expression tag	UNP Q9HCG8
A	117	SER	-	expression tag	UNP Q9HCG8
A	118	MET	-	expression tag	UNP Q9HCG8
B	116	ARG	-	expression tag	UNP Q9HCG8
B	117	SER	-	expression tag	UNP Q9HCG8
B	118	MET	-	expression tag	UNP Q9HCG8
D	116	ARG	-	expression tag	UNP Q9HCG8
D	117	SER	-	expression tag	UNP Q9HCG8
D	118	MET	-	expression tag	UNP Q9HCG8
F	116	ARG	-	expression tag	UNP Q9HCG8
F	117	SER	-	expression tag	UNP Q9HCG8
F	118	MET	-	expression tag	UNP Q9HCG8

- Molecule 2 is a protein called Spliceosome-associated protein CWC27 homolog.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	49	Total	C	N	O	S	0	0	0
			356	222	58	75	1			
2	I	49	Total	C	N	O	S	0	0	0
			349	220	57	71	1			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	49	Total	C	N	O	S	0	0	0
			347	218	56	72	1			
2	G	44	Total	C	N	O	S	0	0	0
			332	208	56	67	1			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	375	ARG	-	expression tag	UNP Q6UX04
C	376	SER	-	expression tag	UNP Q6UX04
C	377	MET	-	expression tag	UNP Q6UX04
I	375	ARG	-	expression tag	UNP Q6UX04
I	376	SER	-	expression tag	UNP Q6UX04
I	377	MET	-	expression tag	UNP Q6UX04
E	375	ARG	-	expression tag	UNP Q6UX04
E	376	SER	-	expression tag	UNP Q6UX04
E	377	MET	-	expression tag	UNP Q6UX04
G	375	ARG	-	expression tag	UNP Q6UX04
G	376	SER	-	expression tag	UNP Q6UX04
G	377	MET	-	expression tag	UNP Q6UX04

- Molecule 3 is a protein called Eukaryotic initiation factor 4A-III.

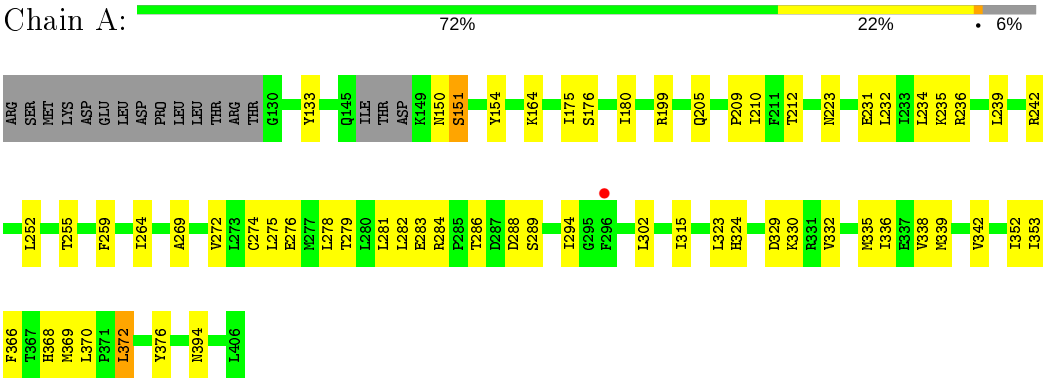
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	H	162	Total	C	N	O	S	0	0	0
			1259	795	209	247	8			
3	J	162	Total	C	N	O	S	0	0	0
			1244	790	205	243	6			
3	K	164	Total	C	N	O	S	0	1	0
			1128	715	192	215	6			
3	L	162	Total	C	N	O	S	0	0	0
			1231	784	198	241	8			

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.

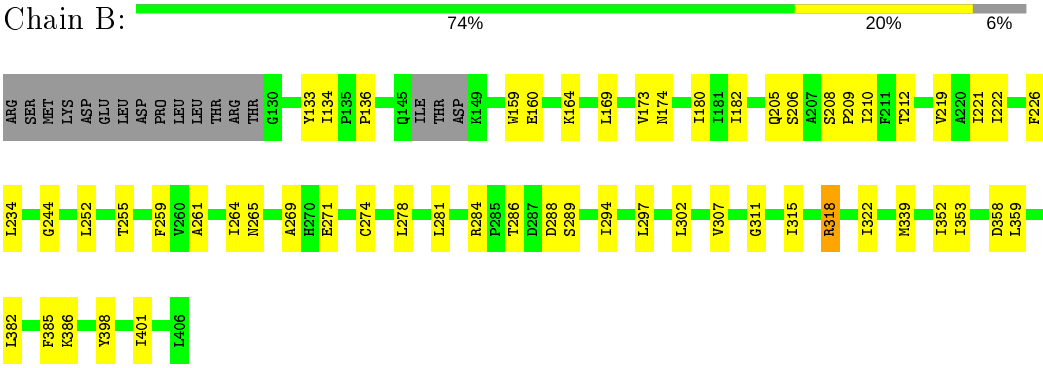
- Molecule 1: Pre-mRNA-splicing factor CWC22 homolog

Chain A:



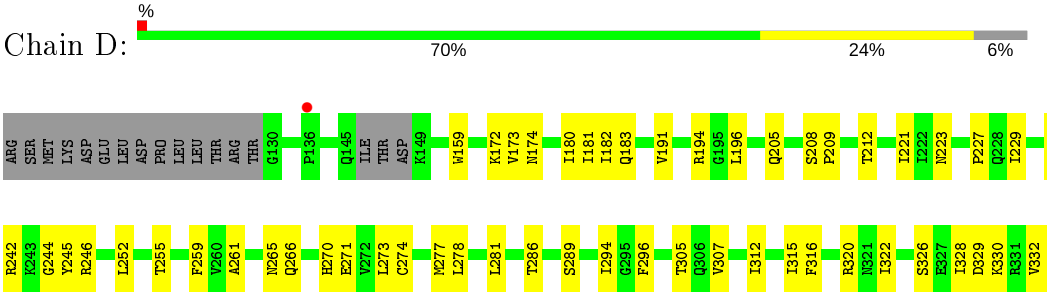
- Molecule 1: Pre-mRNA-splicing factor CWC22 homolog

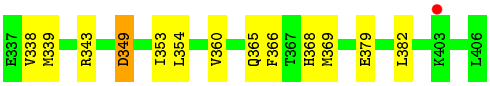
Chain B:



- Molecule 1: Pre-mRNA-splicing factor CWC22 homolog

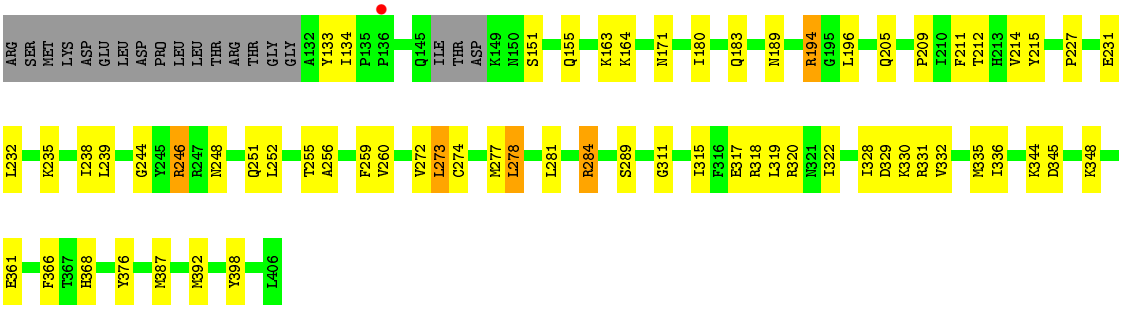
Chain D:





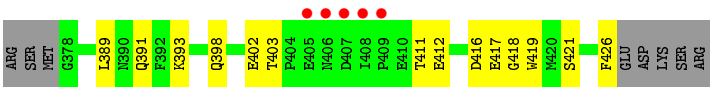
• Molecule 1: Pre-mRNA-splicing factor CWC22 homolog

Chain F:



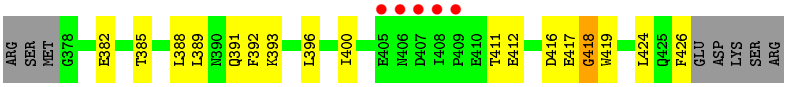
• Molecule 2: Spliceosome-associated protein CWC27 homolog

Chain C:



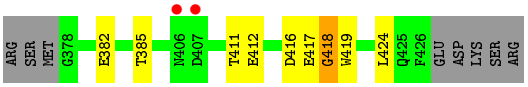
• Molecule 2: Spliceosome-associated protein CWC27 homolog

Chain I:



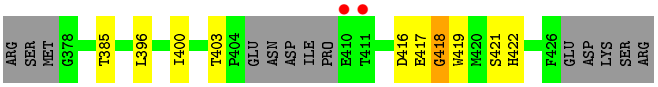
• Molecule 2: Spliceosome-associated protein CWC27 homolog

Chain E:



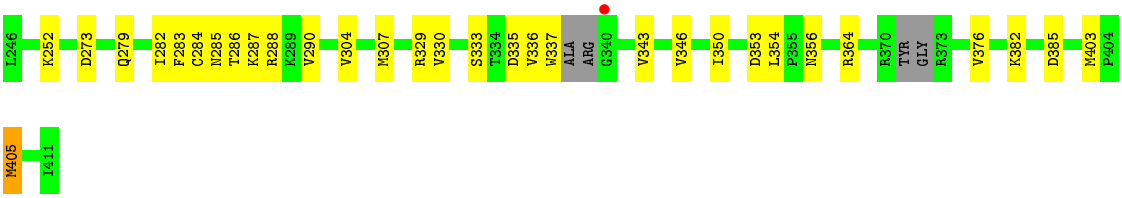
• Molecule 2: Spliceosome-associated protein CWC27 homolog

Chain G:

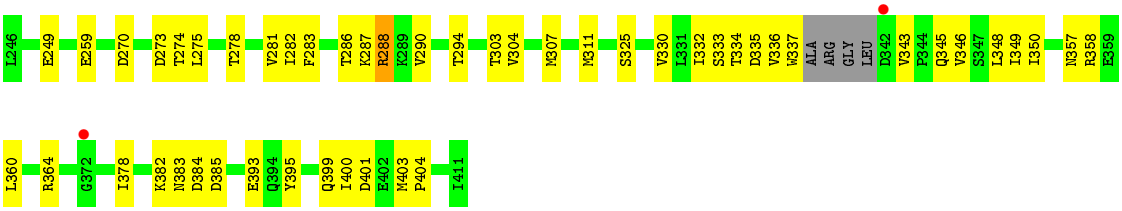


• Molecule 3: Eukaryotic initiation factor 4A-III

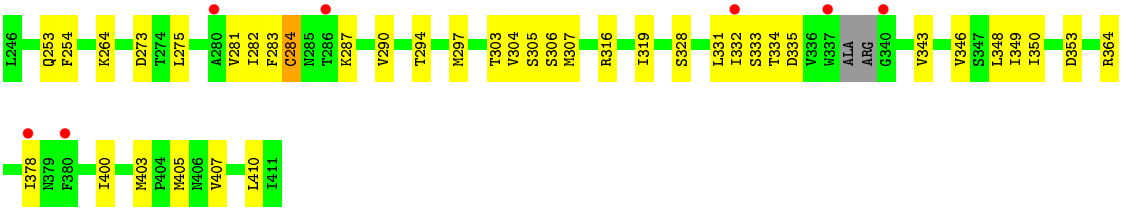
Chain H:



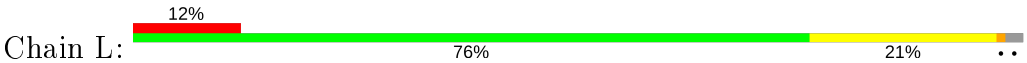
- Molecule 3: Eukaryotic initiation factor 4A-III

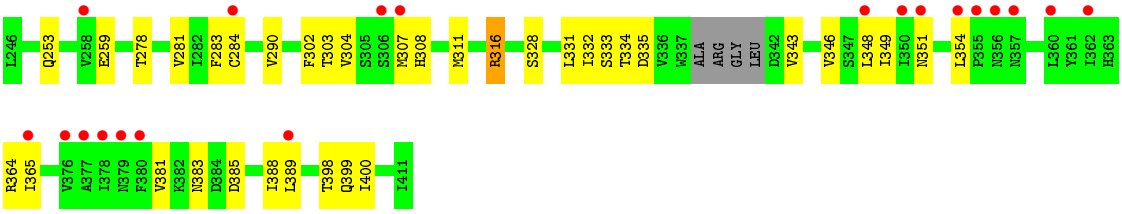


- Molecule 3: Eukaryotic initiation factor 4A-III



- Molecule 3: Eukaryotic initiation factor 4A-III





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	151.25Å 163.63Å 181.13Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	94.69 – 3.19 94.69 – 3.19	Depositor EDS
% Data completeness (in resolution range)	99.3 (94.69-3.19) 99.5 (94.69-3.19)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.48 (at 3.19Å)	Xtriage
Refinement program	PHENIX dev_3758	Depositor
R, R_{free}	0.232 , 0.263 0.234 , 0.261	Depositor DCC
R_{free} test set	3895 reflections (5.20%)	wwPDB-VP
Wilson B-factor (Å^2)	111.5	Xtriage
Anisotropy	0.434	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 78.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	14536	wwPDB-VP
Average B, all atoms (Å^2)	118.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.44% 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	0.26	0/2111	0.48	1/2870 (0.0%)
1	B	0.27	0/2091	0.46	0/2851
1	D	0.27	0/2104	0.49	0/2862
1	F	0.28	0/2131	0.48	0/2894
2	C	0.24	0/362	0.48	0/494
2	E	0.30	0/353	0.47	0/484
2	G	0.27	0/336	0.51	0/454
2	I	0.25	0/355	0.48	0/485
3	H	0.26	0/1279	0.45	0/1734
3	J	0.28	0/1266	0.53	0/1721
3	K	0.27	0/1147	0.46	0/1570
3	L	0.25	0/1253	0.48	0/1706
All	All	0.27	0/14788	0.48	1/20125 (0.0%)

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
2	C	0	1
2	E	0	1
2	G	0	1
2	I	0	1
All	All	0	4

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	372	LEU	CA-CB-CG	5.00	126.81	115.30

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	C	418	GLY	Peptide
2	E	418	GLY	Peptide
2	G	418	GLY	Peptide
2	I	418	GLY	Peptide

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	2074	0	2007	47	0
1	B	2054	0	1965	45	0
1	D	2068	0	1990	48	0
1	F	2094	0	2032	46	0
2	C	356	0	311	10	0
2	E	347	0	299	13	0
2	G	332	0	306	9	0
2	I	349	0	305	16	0
3	H	1259	0	1181	21	0
3	J	1244	0	1151	34	0
3	K	1128	0	969	28	0
3	L	1231	0	1133	30	0
All	All	14536	0	13649	309	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 11.

The worst 5 of 309 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:F:284:ARG:NH2	2:G:421:SER:O	2.12	0.83
2:E:411:THR:HG22	2:E:412:GLU:H	1.43	0.83
1:F:246:ARG:HG3	2:G:418:GLY:H	1.44	0.82
3:L:307:MET:HG3	3:L:316:ARG:HH11	1.46	0.80
1:B:271:GLU:HG2	1:B:307:VAL:HG11	1.63	0.79

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

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	271/291 (93%)	267 (98%)	3 (1%)	1 (0%)	34	69
1	B	271/291 (93%)	266 (98%)	5 (2%)	0	100	100
1	D	271/291 (93%)	268 (99%)	3 (1%)	0	100	100
1	F	269/291 (92%)	262 (97%)	7 (3%)	0	100	100
2	C	47/57 (82%)	40 (85%)	7 (15%)	0	100	100
2	E	47/57 (82%)	40 (85%)	7 (15%)	0	100	100
2	G	40/57 (70%)	36 (90%)	4 (10%)	0	100	100
2	I	47/57 (82%)	40 (85%)	7 (15%)	0	100	100
3	H	156/166 (94%)	151 (97%)	5 (3%)	0	100	100
3	J	158/166 (95%)	149 (94%)	9 (6%)	0	100	100
3	K	161/166 (97%)	154 (96%)	7 (4%)	0	100	100
3	L	158/166 (95%)	152 (96%)	6 (4%)	0	100	100
All	All	1896/2056 (92%)	1825 (96%)	70 (4%)	1 (0%)	51	83

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	151	SER

5.3.2 Protein sidechains

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	211/263 (80%)	207 (98%)	4 (2%)	57 81
1	B	207/263 (79%)	204 (99%)	3 (1%)	67 86
1	D	210/263 (80%)	207 (99%)	3 (1%)	67 86
1	F	219/263 (83%)	210 (96%)	9 (4%)	30 66
2	C	35/52 (67%)	34 (97%)	1 (3%)	42 74
2	E	33/52 (64%)	33 (100%)	0	100 100
2	G	34/52 (65%)	34 (100%)	0	100 100
2	I	33/52 (64%)	33 (100%)	0	100 100
3	H	133/151 (88%)	130 (98%)	3 (2%)	50 78
3	J	128/151 (85%)	122 (95%)	6 (5%)	26 62
3	K	98/151 (65%)	97 (99%)	1 (1%)	76 90
3	L	127/151 (84%)	126 (99%)	1 (1%)	81 93
All	All	1468/1864 (79%)	1437 (98%)	31 (2%)	53 79

5 of 31 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	273	LEU
1	F	318	ARG
3	J	403	MET
1	F	278	LEU
1	F	361	GLU

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

Mol	Chain	Res	Type
1	A	205	GLN
1	A	365	GLN
1	B	205	GLN
1	F	205	GLN
3	L	383	ASN

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 monosaccharides 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	274/291 (94%)	-0.12	1 (0%) 92 89	78, 107, 164, 258	0
1	B	274/291 (94%)	-0.14	0 100 100	77, 107, 175, 307	0
1	D	274/291 (94%)	-0.07	2 (0%) 87 81	66, 106, 173, 286	0
1	F	272/291 (93%)	-0.05	1 (0%) 92 89	67, 97, 180, 216	0
2	C	49/57 (85%)	0.62	5 (10%) 6 4	85, 127, 313, 347	0
2	E	49/57 (85%)	0.06	2 (4%) 37 24	93, 138, 303, 382	0
2	G	44/57 (77%)	-0.04	2 (4%) 33 21	82, 122, 209, 276	0
2	I	49/57 (85%)	0.17	5 (10%) 6 4	87, 123, 241, 318	0
3	H	162/166 (97%)	0.01	1 (0%) 89 83	67, 99, 166, 213	0
3	J	162/166 (97%)	0.15	2 (1%) 79 67	78, 98, 166, 214	0
3	K	164/166 (98%)	0.08	7 (4%) 35 22	82, 140, 188, 224	0
3	L	162/166 (97%)	0.59	20 (12%) 4 2	103, 150, 199, 215	0
All	All	1935/2056 (94%)	0.04	48 (2%) 57 43	66, 112, 191, 382	0

The worst 5 of 48 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	407	ASP	11.3
2	C	406	ASN	10.9
2	C	409	PRO	5.9
2	E	407	ASP	4.7
2	G	410	GLU	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 (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.