wwPDB X-ray Structure Validation Summary Report (i)

May 22, 2020 – 11:07 pm BST

PDB ID	:	2FRX
Title	:	Crystal structure of YebU, a m5C RNA methyltransferase from E.coli
Authors	:	Erlandsen, H.; Nordlund, P.; Hallberg, B.M.; Johnson, K.A.; Ericsson, U.B.
Deposited on	:	2006-01-20
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 following versions of software and data (see references (1)) were used in the production of this report:

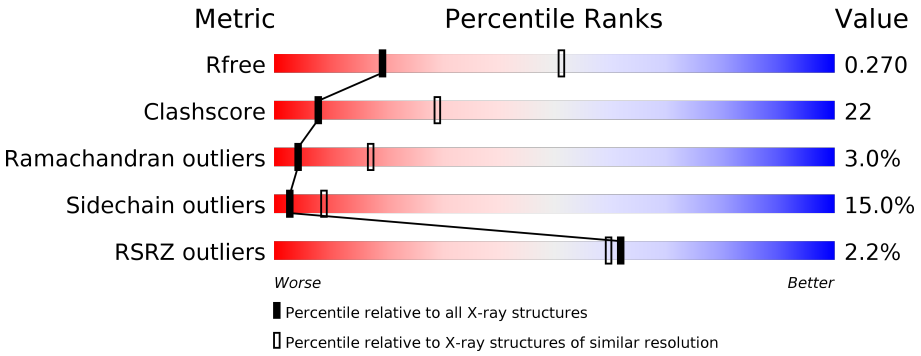
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA,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

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	479	2% 54% 32% 9% 5%
1	B	479	% 54% 32% 8% 5%
1	C	479	2% 50% 36% 8% • 5%
1	D	479	3% 51% 35% 8% • 5%

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 14266 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 Hypothetical protein yebU.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	455	Total	C	N	O	S	Se	0	0	0
			3574	2285	621	654	7	7			
1	B	454	Total	C	N	O	S	Se	0	0	0
			3562	2276	620	652	7	7			
1	C	454	Total	C	N	O	S	Se	0	0	0
			3562	2276	620	652	7	7			
1	D	455	Total	C	N	O	S	Se	0	0	0
			3568	2282	618	654	7	7			

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	MET	MODIFIED RESIDUE	UNP P76273
A	16	MSE	MET	MODIFIED RESIDUE	UNP P76273
A	20	MSE	MET	MODIFIED RESIDUE	UNP P76273
A	105	MSE	MET	MODIFIED RESIDUE	UNP P76273
A	122	MSE	MET	MODIFIED RESIDUE	UNP P76273
A	139	MSE	MET	MODIFIED RESIDUE	UNP P76273
A	187	MSE	MET	MODIFIED RESIDUE	UNP P76273
A	411	MSE	MET	MODIFIED RESIDUE	UNP P76273
B	1	MSE	MET	MODIFIED RESIDUE	UNP P76273
B	16	MSE	MET	MODIFIED RESIDUE	UNP P76273
B	20	MSE	MET	MODIFIED RESIDUE	UNP P76273
B	105	MSE	MET	MODIFIED RESIDUE	UNP P76273
B	122	MSE	MET	MODIFIED RESIDUE	UNP P76273
B	139	MSE	MET	MODIFIED RESIDUE	UNP P76273
B	187	MSE	MET	MODIFIED RESIDUE	UNP P76273
B	411	MSE	MET	MODIFIED RESIDUE	UNP P76273
C	1	MSE	MET	MODIFIED RESIDUE	UNP P76273
C	16	MSE	MET	MODIFIED RESIDUE	UNP P76273
C	20	MSE	MET	MODIFIED RESIDUE	UNP P76273
C	105	MSE	MET	MODIFIED RESIDUE	UNP P76273
C	122	MSE	MET	MODIFIED RESIDUE	UNP P76273

Continued on next page...

Continued from previous page...

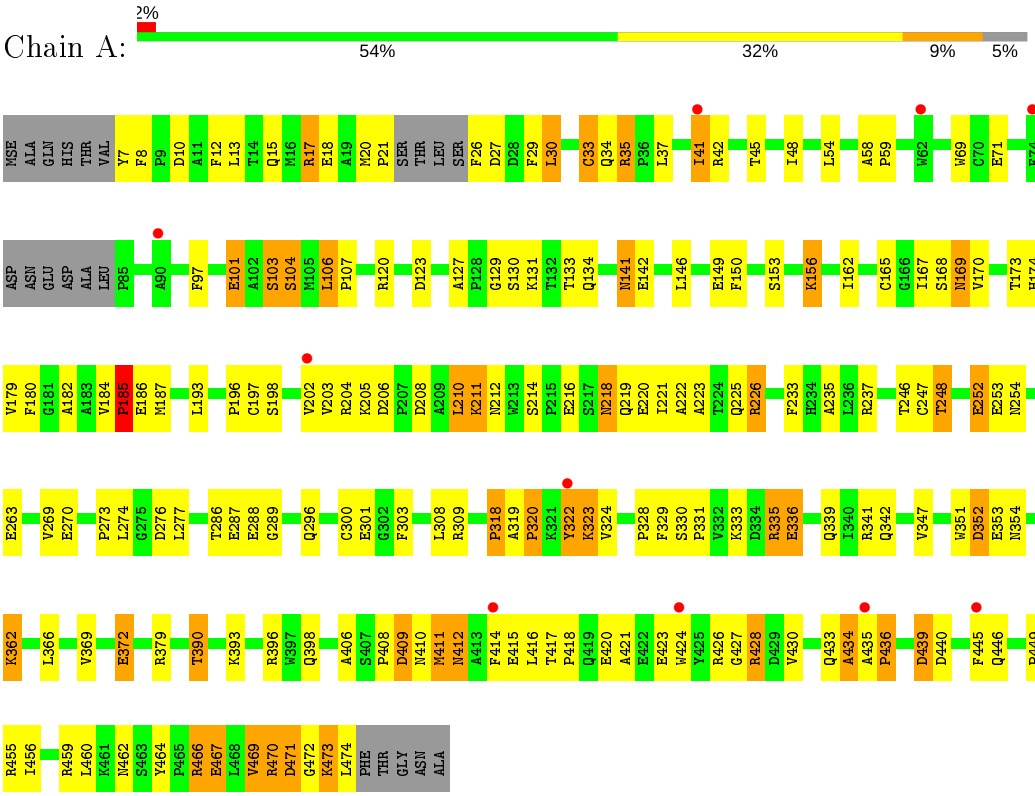
Chain	Residue	Modelled	Actual	Comment	Reference
C	139	MSE	MET	MODIFIED RESIDUE	UNP P76273
C	187	MSE	MET	MODIFIED RESIDUE	UNP P76273
C	411	MSE	MET	MODIFIED RESIDUE	UNP P76273
D	1	MSE	MET	MODIFIED RESIDUE	UNP P76273
D	16	MSE	MET	MODIFIED RESIDUE	UNP P76273
D	20	MSE	MET	MODIFIED RESIDUE	UNP P76273
D	105	MSE	MET	MODIFIED RESIDUE	UNP P76273
D	122	MSE	MET	MODIFIED RESIDUE	UNP P76273
D	139	MSE	MET	MODIFIED RESIDUE	UNP P76273
D	187	MSE	MET	MODIFIED RESIDUE	UNP P76273
D	411	MSE	MET	MODIFIED RESIDUE	UNP P76273

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 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.

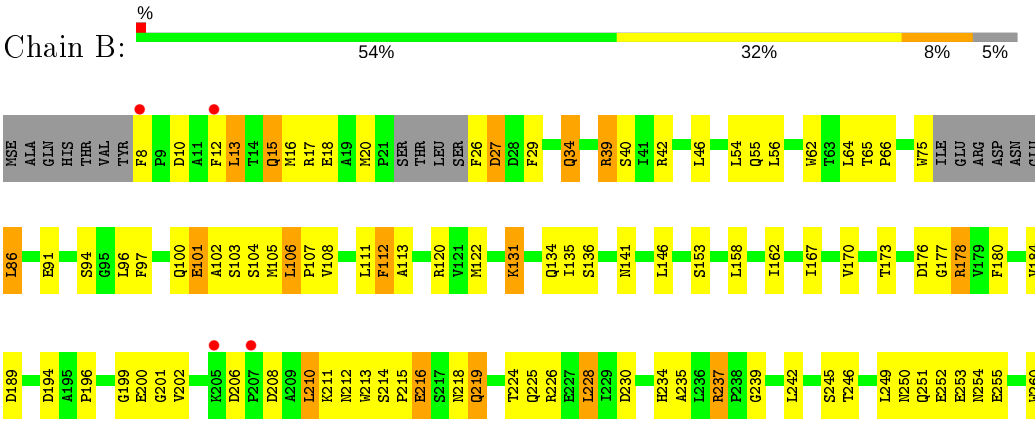
- Molecule 1: Hypothetical protein yebU

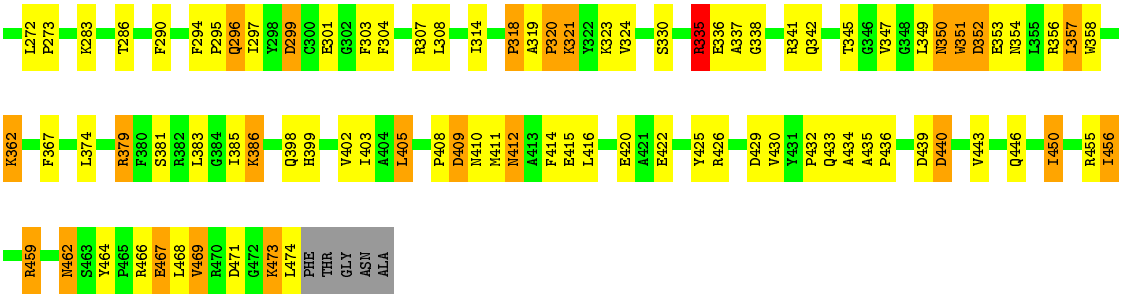
Chain A:



- Molecule 1: Hypothetical protein yebU

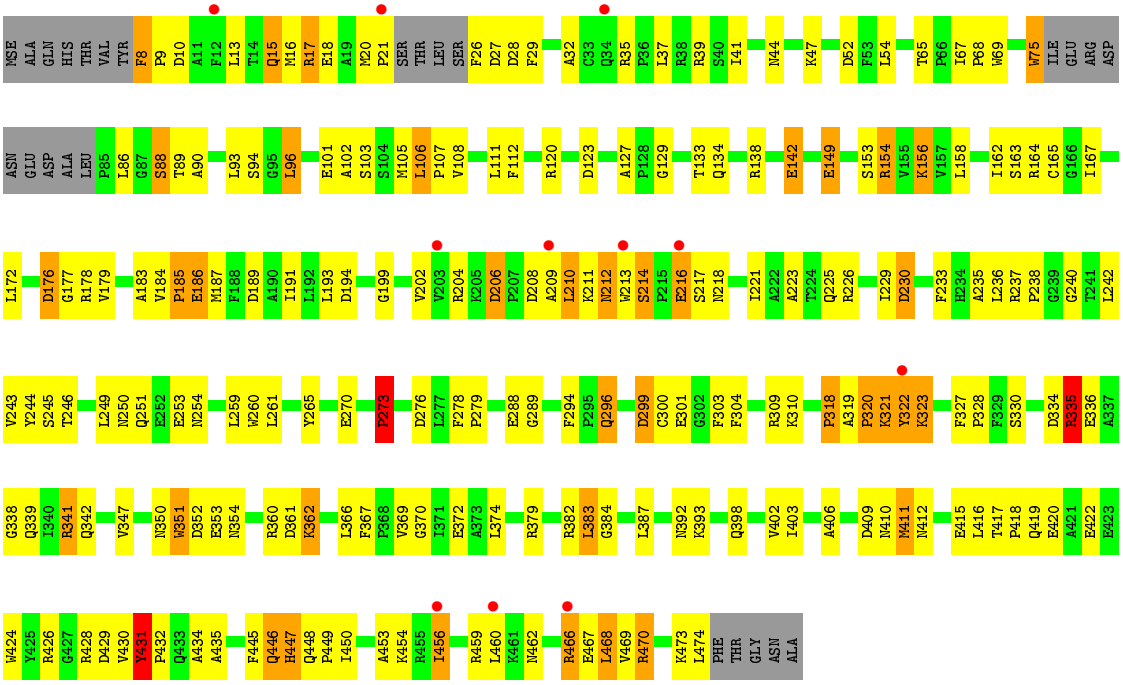
Chain B:





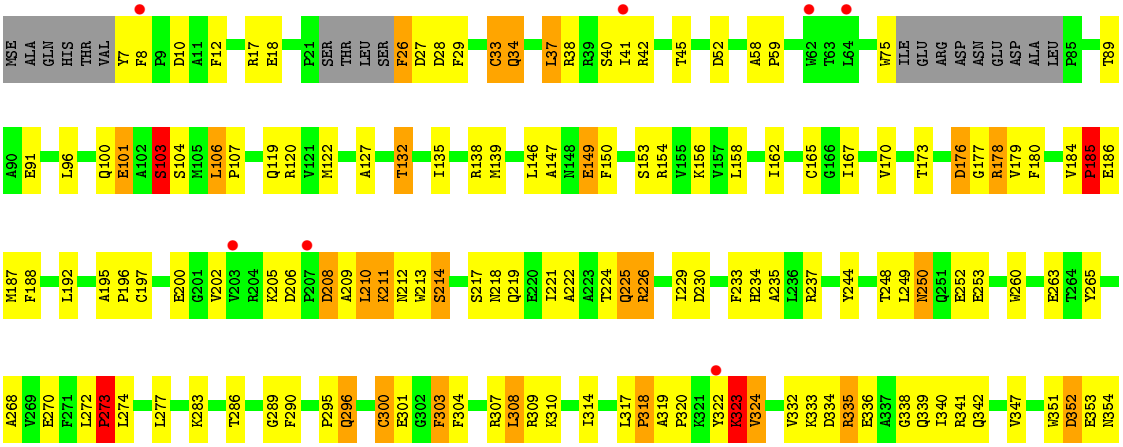
• Molecule 1: Hypothetical protein yebU

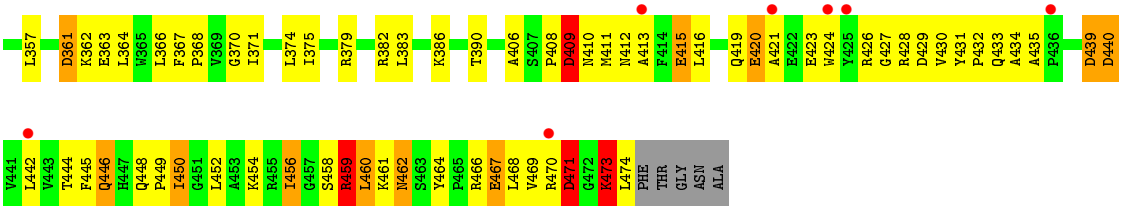
Chain C:



• Molecule 1: Hypothetical protein yebU

Chain D:




L357
D361
K362
E363
L364
W365
L366
F367
P368
V369
G370
I371
L374
I375
R379
R382
L383
K386
T390
A406
S407
P408
D409
N410
M411
N412
A413
F414
E415
L416
Q419
E420
A421
E422
E423
W424
Y425
R426
G427
R428
D429
V430
Y431
P432
Q433
A434
A435
P436
D439
D440
V441
L442
V443
T444
F445
Q446
H447
Q448
P449
I450
G451
L452
A453
K454
R455
I456
G457
S458
R459
L460
K461
N462
S463
Y464
P465
R466
E467
L468
V469
R470
D471
G472
K473
L474
PHE
THR
GLY
ASN
ALA

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	67.71Å 87.13Å 95.05Å 88.33° 76.79° 90.19°	Depositor
Resolution (Å)	29.03 – 2.90 29.03 – 2.90	Depositor EDS
% Data completeness (in resolution range)	97.8 (29.03-2.90) 89.5 (29.03-2.90)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.10	Depositor
$< I/\sigma(I) >$ [1]	1.87 (at 2.90Å)	Xtriage
Refinement program	REFMAC 5.2.0003	Depositor
R, R_{free}	0.231 , 0.282 0.216 , 0.270	Depositor DCC
R_{free} test set	2314 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	67.5	Xtriage
Anisotropy	0.132	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 66.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.093 for -h,k,-l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	14266	wwPDB-VP
Average B, all atoms (Å^2)	65.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.57% 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	0.90	12/3659 (0.3%)	0.97	11/4963 (0.2%)
1	B	0.88	4/3646 (0.1%)	0.98	14/4945 (0.3%)
1	C	1.14	13/3646 (0.4%)	1.16	21/4945 (0.4%)
1	D	1.59	38/3653 (1.0%)	1.17	24/4956 (0.5%)
All	All	1.16	67/14604 (0.5%)	1.07	70/19809 (0.4%)

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
1	B	0	2
1	C	0	1
1	D	0	2
All	All	0	6

The worst 5 of 67 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	335	ARG	NE-CZ	39.64	1.84	1.33
1	D	420	GLU	CD-OE1	38.71	1.68	1.25
1	C	214	SER	CB-OG	33.06	1.85	1.42
1	D	459	ARG	NE-CZ	22.04	1.61	1.33
1	C	428	ARG	CZ-NH1	20.99	1.60	1.33

The worst 5 of 70 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	428	ARG	NE-CZ-NH2	-30.98	104.81	120.30
1	C	428	ARG	NE-CZ-NH1	30.08	135.34	120.30
1	D	17	ARG	NE-CZ-NH2	22.80	131.70	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	459	ARG	NE-CZ-NH2	-19.50	110.55	120.30
1	D	335	ARG	NE-CZ-NH2	-19.42	110.59	120.30

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	318	PRO	Peptide
1	B	318	PRO	Peptide
1	B	323	LYS	Peptide
1	C	318	PRO	Peptide
1	D	318	PRO	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	3574	0	3513	132	0
1	B	3562	0	3504	150	0
1	C	3562	0	3504	160	0
1	D	3568	0	3502	171	0
All	All	14266	0	14023	611	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 22.

The worst 5 of 611 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:D:473:LYS:NZ	1:D:473:LYS:CE	1.68	1.54
1:A:187:MSE:SE	1:A:187:MSE:CE	2.15	1.45
1:A:411:MSE:CE	1:A:411:MSE:SE	2.14	1.45
1:A:454:LYS:NZ	1:A:454:LYS:CE	1.78	1.43
1:C:187:MSE:CE	1:C:187:MSE:SE	2.17	1.42

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	449/479 (94%)	408 (91%)	28 (6%)	13 (3%)	4 18
1	B	448/479 (94%)	401 (90%)	35 (8%)	12 (3%)	5 19
1	C	448/479 (94%)	401 (90%)	33 (7%)	14 (3%)	4 16
1	D	449/479 (94%)	403 (90%)	32 (7%)	14 (3%)	4 16
All	All	1794/1916 (94%)	1613 (90%)	128 (7%)	53 (3%)	4 17

5 of 53 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	185	PRO
1	A	211	LYS
1	A	409	ASP
1	B	185	PRO
1	B	211	LYS

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	373/385 (97%)	310 (83%)	63 (17%)	2 6
1	B	372/385 (97%)	314 (84%)	58 (16%)	2 8
1	C	372/385 (97%)	320 (86%)	52 (14%)	3 10
1	D	372/385 (97%)	322 (87%)	50 (13%)	4 11
All	All	1489/1540 (97%)	1266 (85%)	223 (15%)	3 9

5 of 223 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	345	THR
1	C	18	GLU
1	D	361	ASP
1	B	353	GLU
1	B	443	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 48 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	419	GLN
1	C	218	ASN
1	D	296	GLN
1	B	448	GLN
1	C	15	GLN

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 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

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	448/479 (93%)	0.08	11 (2%)	57	55	52, 65, 70, 77	0
1	B	447/479 (93%)	0.01	4 (0%)	84	84	51, 65, 70, 76	0
1	C	447/479 (93%)	0.02	11 (2%)	57	55	51, 65, 70, 77	0
1	D	448/479 (93%)	0.13	14 (3%)	49	44	52, 65, 70, 76	0
All	All	1790/1916 (93%)	0.06	40 (2%)	62	59	51, 65, 70, 77	0

The worst 5 of 40 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	442	LEU	3.7
1	A	445	PHE	3.4
1	C	213	TRP	3.3
1	C	466	ARG	3.2
1	A	62	TRP	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

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.