



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 6, 2026 – 04:06 PM UTC

PDB ID : 3AV2 / pdb_00003av2
Title : The human nucleosome structure containing the histone variant H3.3
Authors : Tachiwana, H.; Osakabe, A.; Shiga, T.; Miya, M.; Kimura, H.; Kagawa, W.; Kurumizaka, H.
Deposited on : 2011-02-18
Resolution : 2.80 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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity	:	4-5-2 with Phenix2.0
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

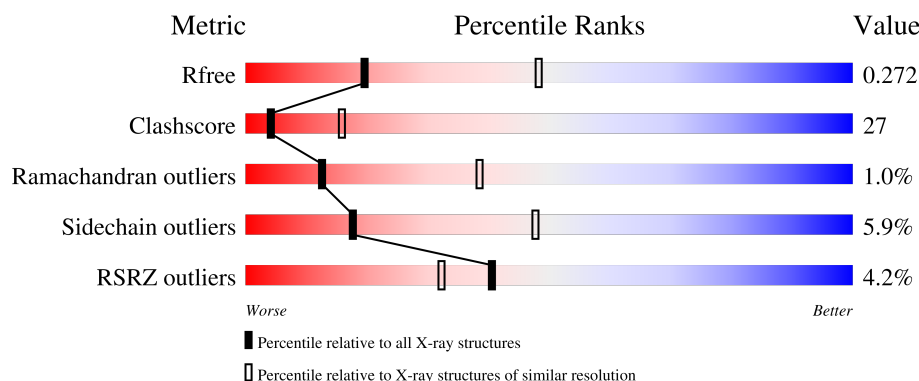
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.80 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}	180053	3866 (2.80-2.80)
Clashscore	190562	4276 (2.80-2.80)
Ramachandran outliers	187476	4196 (2.80-2.80)
Sidechain outliers	187428	4198 (2.80-2.80)
RSRZ outliers	180081	3869 (2.80-2.80)

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 $\leq 5\%$. 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	139	42% 27% • 30%
1	E	139	45% 22% • • 29%
2	B	106	42% 28% • 26%
2	F	106	57% 23% 21%
3	C	133	52% 23% • 21%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	G	133	% 53%23%22%
4	D	129	2% 44%25%27%
4	H	129	% 47%21%29%
5	I	146	12%17% 83%
5	J	146	12%8% 92%

2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 12026 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 Histone H3.3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	97	Total	C	N	O	S	0	0	0
			797	503	155	137	2			
1	E	98	Total	C	N	O	S	0	0	0
			803	506	156	139	2			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-3	GLY	-	expression tag	UNP P84243
A	-2	SER	-	expression tag	UNP P84243
A	-1	HIS	-	expression tag	UNP P84243
E	-3	GLY	-	expression tag	UNP P84243
E	-2	SER	-	expression tag	UNP P84243
E	-1	HIS	-	expression tag	UNP P84243

- Molecule 2 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	78	Total	C	N	O	S	0	0	0
			619	391	120	107	1			
2	F	84	Total	C	N	O	S	0	0	0
			673	424	133	115	1			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-3	GLY	-	expression tag	UNP P62805
B	-2	SER	-	expression tag	UNP P62805
B	-1	HIS	-	expression tag	UNP P62805
F	-3	GLY	-	expression tag	UNP P62805
F	-2	SER	-	expression tag	UNP P62805
F	-1	HIS	-	expression tag	UNP P62805

- Molecule 3 is a protein called Histone H2A type 1-B/E.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	105	Total	C	N	O	0	0	0
			810	511	158	141			
3	G	104	Total	C	N	O	0	0	0
			805	508	157	140			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-3	GLY	-	expression tag	UNP P04908
C	-2	SER	-	expression tag	UNP P04908
C	-1	HIS	-	expression tag	UNP P04908
G	-3	GLY	-	expression tag	UNP P04908
G	-2	SER	-	expression tag	UNP P04908
G	-1	HIS	-	expression tag	UNP P04908

- Molecule 4 is a protein called Histone H2B type 1-J.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	94	Total	C	N	O	S	0	0	0
			736	462	134	138	2			
4	H	92	Total	C	N	O	S	0	0	0
			719	453	129	135	2			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	-3	GLY	-	expression tag	UNP P06899
D	-2	SER	-	expression tag	UNP P06899
D	-1	HIS	-	expression tag	UNP P06899
H	-3	GLY	-	expression tag	UNP P06899
H	-2	SER	-	expression tag	UNP P06899
H	-1	HIS	-	expression tag	UNP P06899

- Molecule 5 is a DNA chain called 146-MER DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	I	146	Total	C	N	O	P	0	0	0
			2990	1431	540	874	145			
5	J	146	Total	C	N	O	P	0	0	0
			2990	1431	540	874	145			

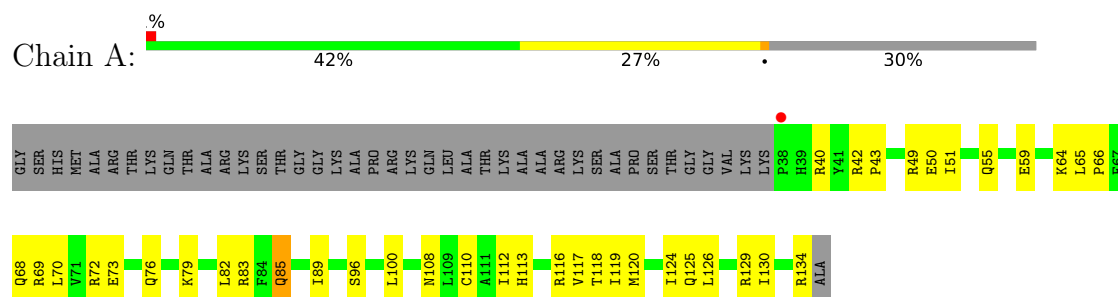
- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	6	Total O 6 6	0	0
6	B	5	Total O 5 5	0	0
6	C	10	Total O 10 10	0	0
6	D	4	Total O 4 4	0	0
6	E	18	Total O 18 18	0	0
6	F	14	Total O 14 14	0	0
6	G	5	Total O 5 5	0	0
6	H	1	Total O 1 1	0	0
6	I	11	Total O 11 11	0	0
6	J	10	Total O 10 10	0	0

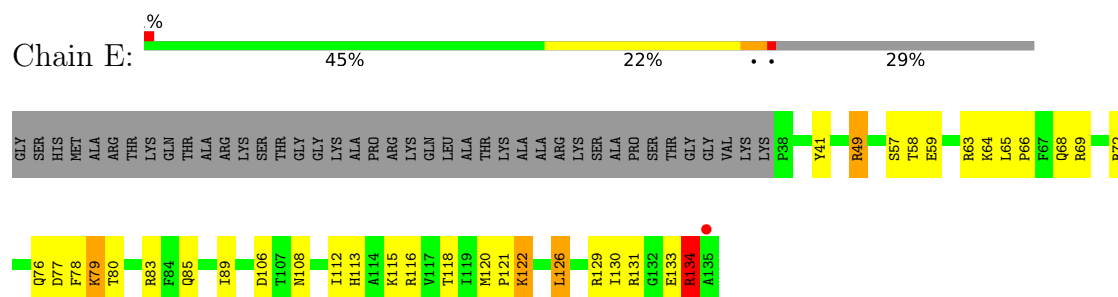
3 Residue-property plots

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.

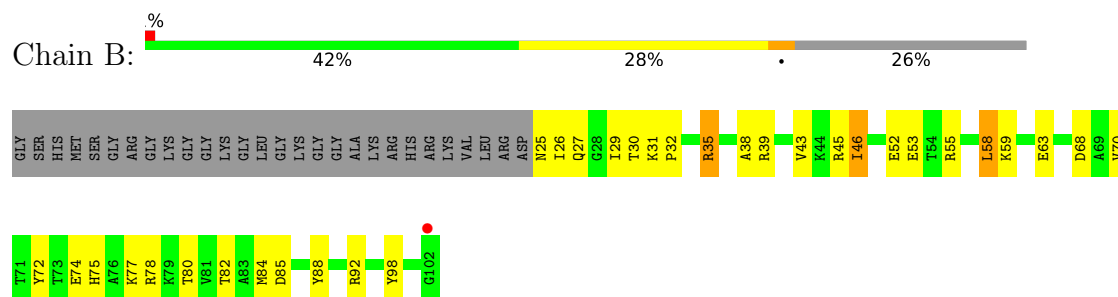
• Molecule 1: Histone H3.3



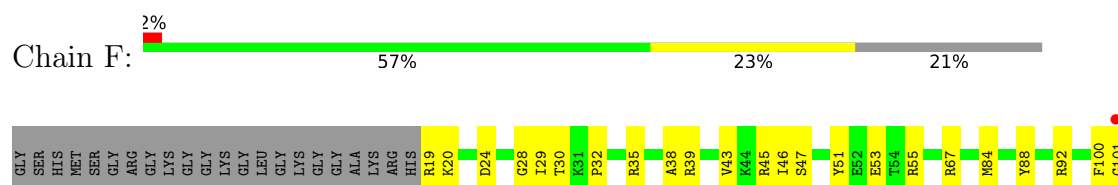
• Molecule 1: Histone H3.3



• Molecule 2: Histone H4



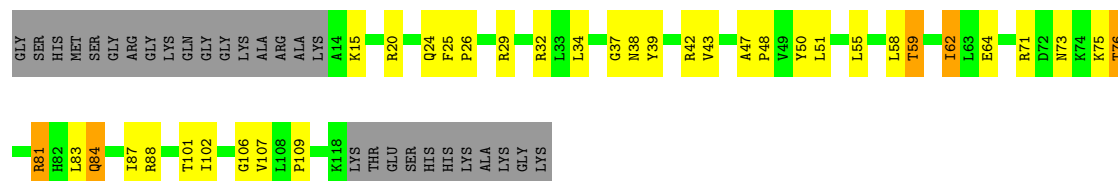
• Molecule 2: Histone H4



G102

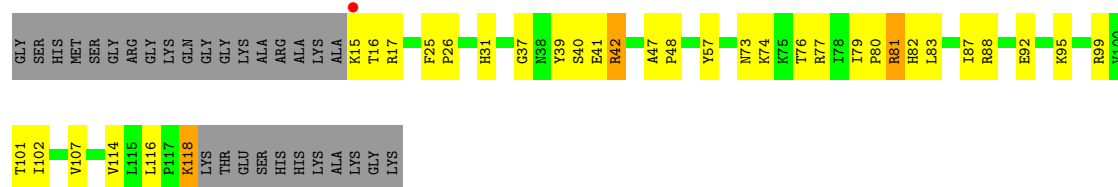
• Molecule 3: Histone H2A type 1-B/E

Chain C: 52% 23% 21%



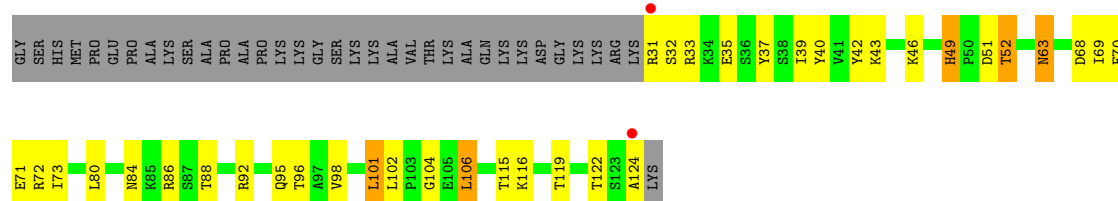
• Molecule 3: Histone H2A type 1-B/E

Chain G: 53% 23% 22%



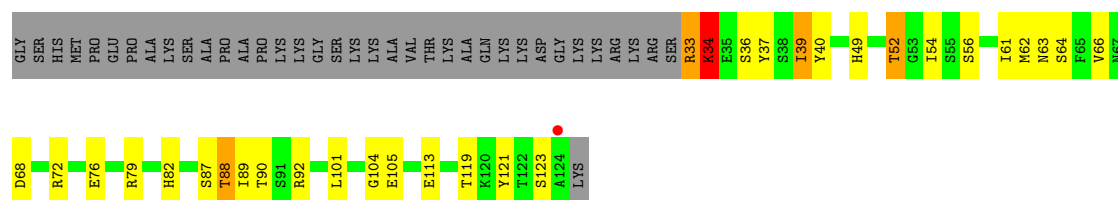
• Molecule 4: Histone H2B type 1-J

Chain D: 44% 25% 27%



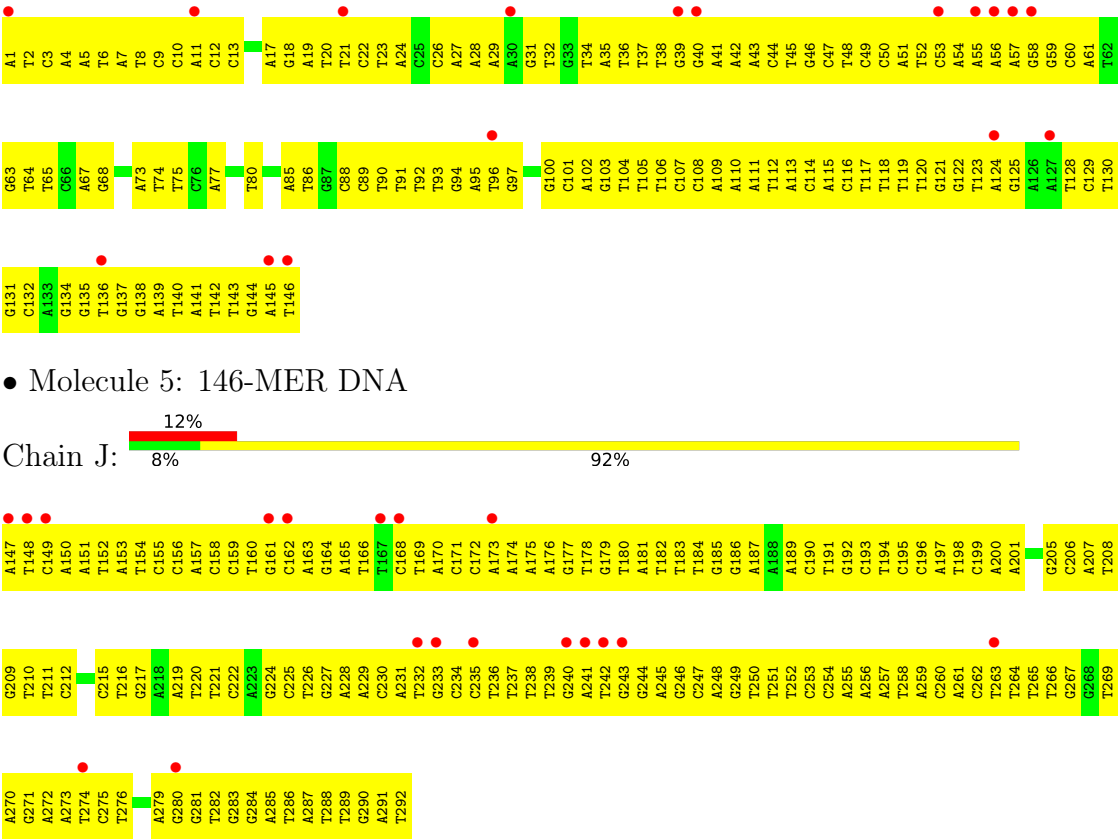
• Molecule 4: Histone H2B type 1-J

Chain H: 47% 21% 29%



• Molecule 5: 146-MER DNA

Chain I: 12% 17% 83%



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	105.84Å 109.69Å 181.18Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.80 50.00 – 2.80	Depositor EDS
% Data completeness (in resolution range)	(Not available) (50.00-2.80) 99.8 (50.00-2.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.07	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.38 (at 2.81Å)	Xtriage
Refinement program	CNS 1.2.1	Depositor
R, R_{free}	0.230 , 0.272 0.230 , 0.272	Depositor DCC
R_{free} test set	2674 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å ²)	55.7	Xtriage
Anisotropy	0.426	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 66.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	0.025 for k,h,-l	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	12026	wwPDB-VP
Average B, all atoms (Å ²)	75.0	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.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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	0.54	0/809	1.00	2/1085 (0.2%)
1	E	0.60	0/815	0.99	2/1092 (0.2%)
2	B	0.54	0/626	1.03	4/837 (0.5%)
2	F	0.64	0/680	1.05	3/908 (0.3%)
3	C	0.58	0/820	1.03	1/1107 (0.1%)
3	G	0.57	0/815	0.99	4/1100 (0.4%)
4	D	0.59	0/747	1.06	5/1004 (0.5%)
4	H	0.56	0/730	0.92	3/982 (0.3%)
5	I	0.31	0/3354	0.76	0/5175
5	J	0.31	0/3354	0.76	0/5175
All	All	0.46	0/12750	0.88	24/18465 (0.1%)

There are no bond length outliers.

The worst 5 of 24 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	D	49	HIS	CA-C-N	7.83	128.56	119.47
4	D	49	HIS	C-N-CA	7.83	128.56	119.47
2	B	30	THR	N-CA-C	7.71	120.71	110.53
1	A	64	LYS	N-CA-C	6.24	118.63	111.02
3	C	43	VAL	N-CA-C	6.17	116.76	107.75

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	797	0	835	36	0
1	E	803	0	840	33	0
2	B	619	0	659	40	0
2	F	673	0	722	18	0
3	C	810	0	866	36	0
3	G	805	0	861	35	0
4	D	736	0	758	40	0
4	H	719	0	740	30	0
5	I	2990	0	1652	182	0
5	J	2990	0	1652	215	0
6	A	6	0	0	0	0
6	B	5	0	0	1	0
6	C	10	0	0	1	0
6	D	4	0	0	1	0
6	E	18	0	0	1	0
6	F	14	0	0	0	0
6	G	5	0	0	0	0
6	H	1	0	0	0	0
6	I	11	0	0	0	0
6	J	10	0	0	3	0
All	All	12026	0	9585	587	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 27.

The worst 5 of 587 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:I:36:DT:H2''	5:I:37:DT:H5''	1.22	1.14
1:E:134:ARG:HH11	1:E:134:ARG:HB3	1.12	1.14
5:I:51:DA:H2''	5:I:52:DT:H5''	1.28	1.12
5:I:20:DT:H2''	5:I:21:DT:H5''	1.32	1.12
5:I:52:DT:H2''	5:I:53:DC:H5'	1.34	1.04

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	95/139 (68%)	94 (99%)	1 (1%)	0	100	100
1	E	96/139 (69%)	91 (95%)	4 (4%)	1 (1%)	12	38
2	B	76/106 (72%)	74 (97%)	2 (3%)	0	100	100
2	F	82/106 (77%)	78 (95%)	3 (4%)	1 (1%)	10	34
3	C	103/133 (77%)	98 (95%)	5 (5%)	0	100	100
3	G	102/133 (77%)	96 (94%)	6 (6%)	0	100	100
4	D	92/129 (71%)	86 (94%)	5 (5%)	1 (1%)	11	36
4	H	90/129 (70%)	84 (93%)	2 (2%)	4 (4%)	2	7
All	All	736/1014 (73%)	701 (95%)	28 (4%)	7 (1%)	12	38

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	H	34	LYS
4	H	104	GLY
4	H	123	SER
4	D	104	GLY
1	E	134	ARG

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	83/112 (74%)	79 (95%)	4 (5%)	23	56

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	E	83/112 (74%)	78 (94%)	5 (6%)	17	47
2	B	63/81 (78%)	62 (98%)	1 (2%)	55	83
2	F	69/81 (85%)	67 (97%)	2 (3%)	37	73
3	C	83/102 (81%)	75 (90%)	8 (10%)	8	26
3	G	83/102 (81%)	77 (93%)	6 (7%)	13	39
4	D	80/107 (75%)	75 (94%)	5 (6%)	16	45
4	H	78/107 (73%)	72 (92%)	6 (8%)	12	36
All	All	622/804 (77%)	585 (94%)	37 (6%)	18	48

5 of 37 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	G	101	THR
4	H	88	THR
3	G	114	VAL
4	H	34	LYS
3	C	109	PRO

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 20 such sidechains are listed below:

Mol	Chain	Res	Type
2	F	93	GLN
3	G	110	ASN
4	H	63	ASN
4	H	47	GLN
3	C	73	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 [i](#)

There are no oligosaccharides 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 [i](#)

6.1 Protein, DNA and RNA chains [i](#)

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, 95th 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	97/139 (69%)	-0.19	1 (1%) 79 72	32, 46, 68, 93	0
1	E	98/139 (70%)	-0.39	1 (1%) 79 72	25, 38, 61, 77	0
2	B	78/106 (73%)	-0.36	1 (1%) 75 66	31, 46, 62, 71	0
2	F	84/106 (79%)	-0.37	2 (2%) 59 49	23, 38, 53, 84	0
3	C	105/133 (78%)	-0.31	0 100 100	24, 43, 67, 82	0
3	G	104/133 (78%)	-0.26	1 (0%) 79 72	33, 49, 73, 86	0
4	D	94/129 (72%)	-0.13	2 (2%) 63 54	29, 46, 71, 122	0
4	H	92/129 (71%)	-0.12	1 (1%) 78 70	34, 49, 74, 89	0
5	I	146/146 (100%)	1.01	17 (11%) 9 7	47, 104, 143, 168	0
5	J	146/146 (100%)	1.12	18 (12%) 8 6	54, 104, 145, 163	0
All	All	1044/1306 (79%)	0.11	44 (4%) 40 32	23, 50, 132, 168	0

The worst 5 of 44 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	D	124	ALA	4.2
2	B	102	GLY	4.1
1	A	38	PRO	4.0
5	I	146	DT	3.9
4	H	124	ALA	3.9

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

There are no oligosaccharides 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.