wwPDB X-ray Structure Validation Summary Report (i)

Sep 2, 2023 – 04:10 PM EDT

PDB ID	:	3REJ
Title	:	2.55 Angstrom Crystal Structure of the Nucleosome Core Particle Assembled with a 146 bp Alpha-Satellite DNA (NCP146b)
Authors	:	Wu, B.; Davey, C.A.
Deposited on	:	2011-04-04
Resolution	:	2.55 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 (1)) were used in the production of this report:

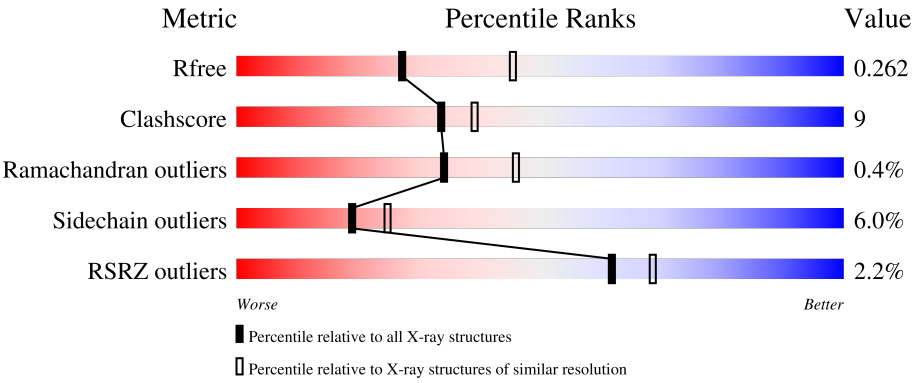
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 2.55 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	1284 (2.56-2.52)
Clashscore	141614	1332 (2.56-2.52)
Ramachandran outliers	138981	1315 (2.56-2.52)
Sidechain outliers	138945	1315 (2.56-2.52)
RSRZ outliers	127900	1272 (2.56-2.52)

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	135	60% 11% 27%
1	E	135	63% 9% 28%
2	B	102	66% 16% 19%
2	F	102	68% 8% 24%
3	C	129	55% 22% 20%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	G	129	2% 60% 18% 19%
4	D	122	7% 57% 20% 19%
4	H	122	% 61% 13% 23%
5	I	146	58% 35% 8%
5	J	146	% 60% 35% 5%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
7	SO4	D	1101	-	-	X	-

2 Entry composition (i)

There are 8 unique types of molecules in this entry. The entry contains 12413 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 H3.2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	98	Total	C	N	O	S	0	0	0
			808	509	156	140	3			
1	E	97	Total	C	N	O	S	0	0	0
			801	504	155	139	3			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	102	ALA	GLY	variant	UNP P84233
E	102	ALA	GLY	variant	UNP P84233

- Molecule 2 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	83	Total	C	N	O	S	0	0	0
			662	418	129	114	1			
2	F	78	Total	C	N	O	S	0	0	0
			619	391	120	107	1			

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

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	103	Total	C	N	O	0	0	0
			795	501	155	139			
3	G	105	Total	C	N	O	0	0	0
			809	510	158	141			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	99	ARG	GLY	variant	UNP P06897
C	123	SER	ALA	variant	UNP P06897

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
G	99	ARG	GLY	variant	UNP P06897
G	123	SER	ALA	variant	UNP P06897

- Molecule 4 is a protein called Histone H2B 1.1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	99	Total	C	N	O	S	0	0	0
			785	493	146	144	2			
4	H	94	Total	C	N	O	S	0	0	0
			736	463	132	139	2			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	29	THR	SER	variant	UNP P02281
H	29	THR	SER	variant	UNP P02281

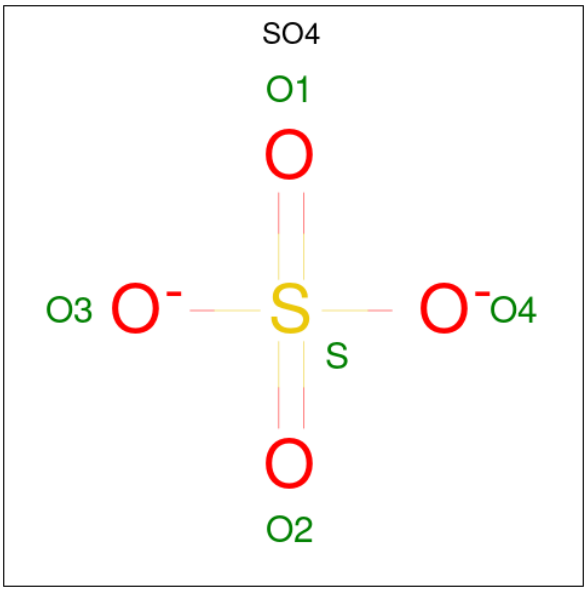
- Molecule 5 is a DNA chain called DNA (146-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	I	146	Total	C	N	O	P	0	0	0
			2990	1430	541	874	145			
5	J	146	Total	C	N	O	P	0	0	0
			2990	1430	541	874	145			

- Molecule 6 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	1	Total	Mn	0	0
			1	1		
6	D	1	Total	Mn	0	0
			1	1		
6	I	4	Total	Mn	0	0
			4	4		
6	J	7	Total	Mn	0	0
			7	7		

- Molecule 7 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	C	1	Total O S 5 4 1	0	0
7	D	1	Total O S 5 4 1	0	0
7	G	1	Total O S 5 4 1	0	0

- Molecule 8 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
8	A	59	Total O 59 59	0	0
8	B	40	Total O 40 40	0	0
8	C	34	Total O 34 34	0	0
8	D	14	Total O 14 14	0	0
8	E	32	Total O 32 32	0	0
8	F	24	Total O 24 24	0	0
8	G	47	Total O 47 47	0	0
8	H	29	Total O 29 29	0	0
8	I	50	Total O 50 50	0	0

Continued on next page...

Continued from previous page...

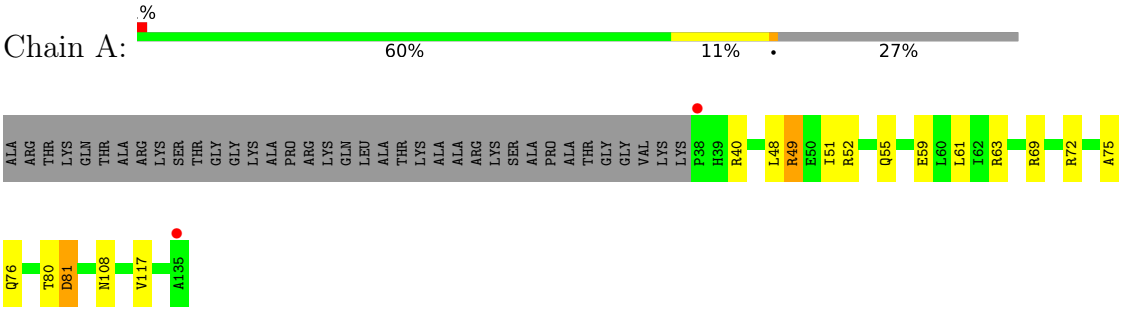
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	J	61	Total	O	0	0
			61	61		

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.

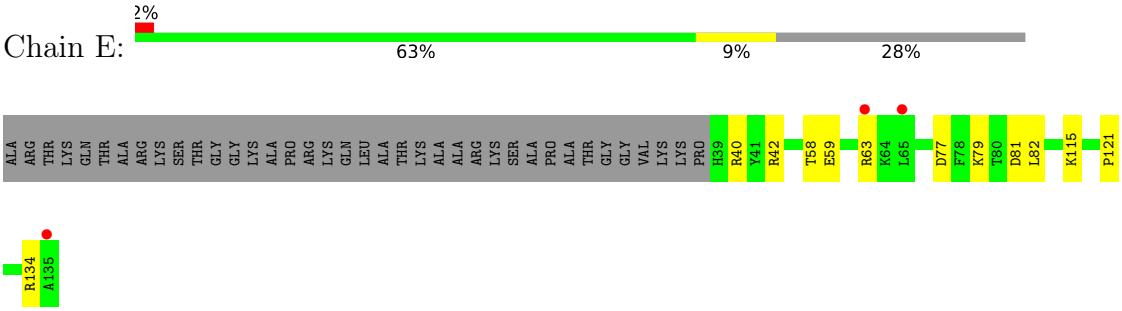
• Molecule 1: Histone H3.2

Chain A:



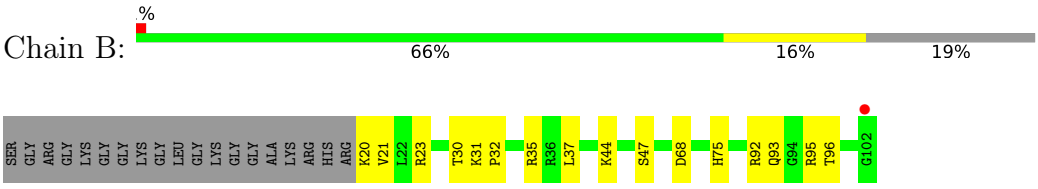
• Molecule 1: Histone H3.2

Chain E:



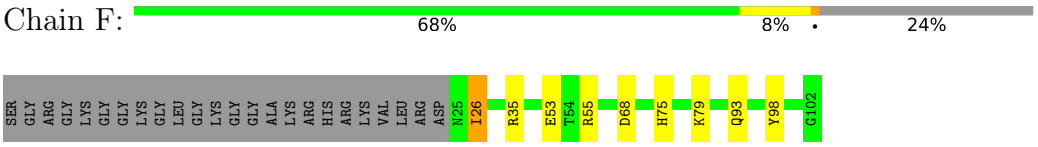
• Molecule 2: Histone H4

Chain B:



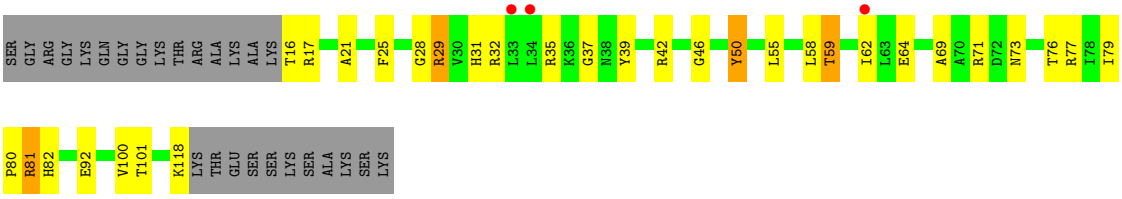
• Molecule 2: Histone H4

Chain F:



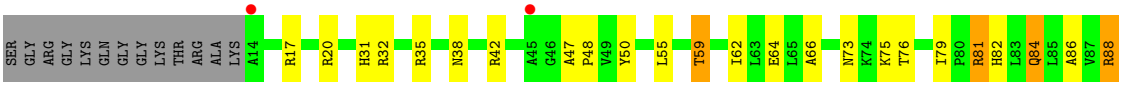
• Molecule 3: Histone H2A type 1

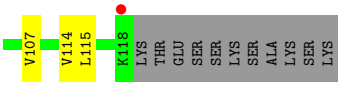
Chain C:



• Molecule 3: Histone H2A type 1

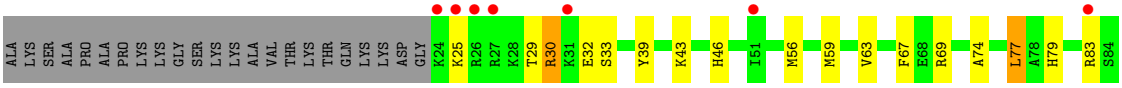
Chain G:

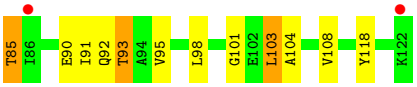




• Molecule 4: Histone H2B 1.1

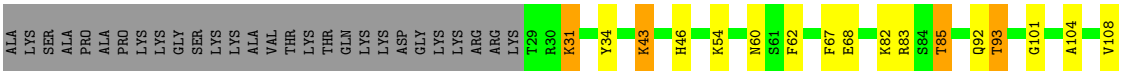
Chain D:





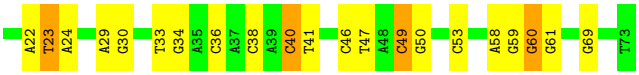
• Molecule 4: Histone H2B 1.1

Chain H:



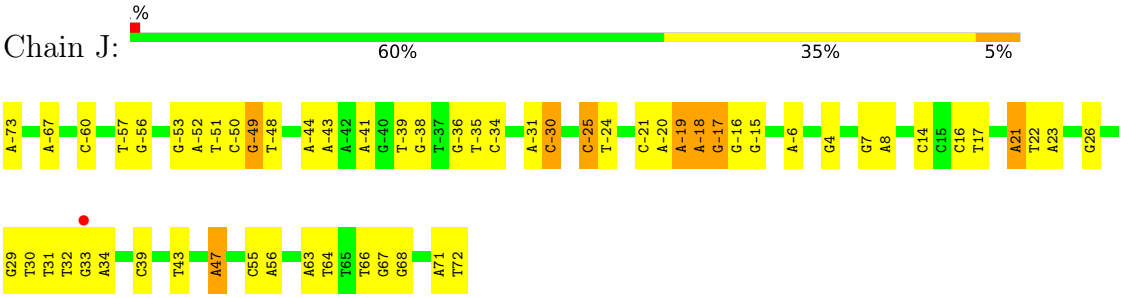
• Molecule 5: DNA (146-MER)

Chain I:



• Molecule 5: DNA (146-MER)

Chain J:


%
60%
35%
5%
A-73 A-67 C-60 T-57 G-56 G-53 A-52 T-51 C-50 G-49 T-48 A-44 A-43 A-42 A-41 G-40 T-39 G-38 T-37 G-36 T-35 C-34 A-31 C-30 C-25 T-24 C-21 A-20 A-19 A-18 G-17 G-16 G-15 A-6 G4 G7 A8 C14 C15 C16 T17 A21 T22 A23 G26
G29 T30 T31 T32 G33 A34 C39 T43 A47 C55 A56 A63 T64 T65 T66 G67 G68 A71 T72

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	106.07Å 109.37Å 176.47Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	92.85 – 2.55 92.97 – 2.55	Depositor EDS
% Data completeness (in resolution range)	91.4 (92.85-2.55) 91.4 (92.97-2.55)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.33 (at 2.55Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.234 , 0.262 0.228 , 0.262	Depositor DCC
R_{free} test set	1253 reflections (2.03%)	wwPDB-VP
Wilson B-factor (Å^2)	54.5	Xtriage
Anisotropy	0.443	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 54.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.021 for k,h,-l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	12413	wwPDB-VP
Average B, all atoms (Å^2)	70.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.90% 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: MN, SO4

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	0.60	0/820	0.68	0/1099
1	E	0.48	0/812	0.63	0/1088
2	B	0.62	0/669	0.73	0/894
2	F	0.53	0/626	0.61	0/837
3	C	0.46	0/805	0.62	0/1088
3	G	0.60	0/819	0.71	0/1106
4	D	0.57	0/796	0.65	1/1065 (0.1%)
4	H	0.58	0/747	0.67	0/1004
5	I	0.71	0/3354	1.38	28/5175 (0.5%)
5	J	0.72	0/3354	1.42	27/5175 (0.5%)
All	All	0.64	0/12802	1.13	56/18531 (0.3%)

There are no bond length outliers.

The worst 5 of 56 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	J	30	DT	O4'-C1'-N1	8.67	114.07	108.00
5	J	-30	DC	O4'-C1'-N1	8.24	113.77	108.00
5	J	43	DT	O4'-C1'-N1	7.96	113.57	108.00
5	J	4	DG	O4'-C1'-N9	-7.64	102.65	108.00
5	J	47	DA	O4'-C1'-N9	-7.57	102.70	108.00

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	808	0	846	15	0
1	E	801	0	838	7	0
2	B	662	0	709	16	0
2	F	619	0	659	10	0
3	C	795	0	846	32	0
3	G	809	0	864	31	0
4	D	785	0	825	32	0
4	H	736	0	760	19	0
5	I	2990	0	1651	38	0
5	J	2990	0	1651	42	0
6	A	1	0	0	0	1
6	D	1	0	0	0	0
6	I	4	0	0	0	0
6	J	7	0	0	0	0
7	C	5	0	0	1	0
7	D	5	0	0	2	0
7	G	5	0	0	0	0
8	A	59	0	0	6	0
8	B	40	0	0	3	0
8	C	34	0	0	1	0
8	D	14	0	0	0	0
8	E	32	0	0	2	0
8	F	24	0	0	0	0
8	G	47	0	0	1	0
8	H	29	0	0	2	1
8	I	50	0	0	1	0
8	J	61	0	0	1	0
All	All	12413	0	9649	182	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 9.

The worst 5 of 182 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)
2:F:75:HIS:CD2	4:H:93:THR:HG21	1.78	1.17
2:F:75:HIS:HD2	4:H:93:THR:HG21	1.09	1.14
2:B:75:HIS:CD2	4:D:93:THR:HG21	1.97	0.99
2:B:75:HIS:HD2	4:D:93:THR:HG21	1.28	0.98

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:C:71:ARG:NH2	7:D:1101:SO4:O3	1.97	0.95

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)
6:A:1001:MN:MN	8:H:439:HOH:O[3_545]	1.70	0.50

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	96/135 (71%)	96 (100%)	0	0	100	100
1	E	95/135 (70%)	90 (95%)	5 (5%)	0	100	100
2	B	81/102 (79%)	78 (96%)	3 (4%)	0	100	100
2	F	76/102 (74%)	74 (97%)	2 (3%)	0	100	100
3	C	101/129 (78%)	98 (97%)	3 (3%)	0	100	100
3	G	103/129 (80%)	99 (96%)	4 (4%)	0	100	100
4	D	97/122 (80%)	92 (95%)	3 (3%)	2 (2%)	7	7
4	H	92/122 (75%)	90 (98%)	1 (1%)	1 (1%)	14	19
All	All	741/976 (76%)	717 (97%)	21 (3%)	3 (0%)	34	46

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	D	101	GLY
4	D	30	ARG
4	H	101	GLY

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	85/110 (77%)	79 (93%)	6 (7%)	14	19
1	E	84/110 (76%)	79 (94%)	5 (6%)	19	25
2	B	68/78 (87%)	67 (98%)	1 (2%)	65	77
2	F	63/78 (81%)	62 (98%)	1 (2%)	62	77
3	C	82/101 (81%)	76 (93%)	6 (7%)	14	18
3	G	83/101 (82%)	77 (93%)	6 (7%)	14	18
4	D	85/102 (83%)	80 (94%)	5 (6%)	19	25
4	H	80/102 (78%)	72 (90%)	8 (10%)	7	8
All	All	630/782 (81%)	592 (94%)	38 (6%)	19	25

5 of 38 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	G	88	ARG
4	H	85	THR
3	G	114	VAL
4	H	68	GLU
4	H	109	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
3	G	38	ASN
3	G	84	GLN
4	H	92	GLN
4	H	46	HIS
4	D	79	HIS

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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 16 ligands modelled in this entry, 13 are monoatomic - leaving 3 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
7	SO4	D	1101	-	4,4,4	0.48	0	6,6,6	0.55	0
7	SO4	C	1102	-	4,4,4	0.48	0	6,6,6	0.69	0
7	SO4	G	1103	-	4,4,4	0.46	0	6,6,6	0.68	0

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.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
7	D	1101	SO4	2	0
7	C	1102	SO4	1	0

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

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	98/135 (72%)	0.19	2 (2%) 65 72	25, 37, 58, 70	0
1	E	97/135 (71%)	0.28	3 (3%) 49 56	35, 51, 71, 76	0
2	B	83/102 (81%)	0.01	1 (1%) 79 84	26, 35, 43, 45	0
2	F	78/102 (76%)	0.10	0 100 100	35, 45, 57, 60	0
3	C	103/129 (79%)	0.21	3 (2%) 51 59	39, 50, 68, 76	0
3	G	105/129 (81%)	0.21	3 (2%) 51 59	23, 40, 58, 69	0
4	D	99/122 (81%)	0.62	9 (9%) 9 11	37, 51, 94, 107	0
4	H	94/122 (77%)	0.27	1 (1%) 80 85	29, 41, 61, 78	0
5	I	146/146 (100%)	-0.26	0 100 100	51, 100, 132, 157	0
5	J	146/146 (100%)	-0.30	1 (0%) 87 90	48, 94, 125, 135	0
All	All	1049/1268 (82%)	0.10	23 (2%) 62 68	23, 50, 114, 157	0

The worst 5 of 23 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	D	24	LYS	8.1
3	G	118	LYS	7.9
4	D	27	ARG	5.9
1	A	135	ALA	4.5
4	D	26	ARG	4.3

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)

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
6	MN	J	1012	1/1	0.45	0.23	123,123,123,123	0
6	MN	I	1004	1/1	0.54	0.24	144,144,144,144	0
6	MN	J	1011	1/1	0.59	0.12	135,135,135,135	0
6	MN	I	1006	1/1	0.82	0.12	85,85,85,85	0
6	MN	J	1010	1/1	0.82	0.21	111,111,111,111	0
6	MN	J	1013	1/1	0.82	0.14	139,139,139,139	0
6	MN	I	1005	1/1	0.85	0.23	96,96,96,96	0
6	MN	J	1009	1/1	0.86	0.08	121,121,121,121	0
6	MN	J	1002	1/1	0.91	0.24	110,110,110,110	0
6	MN	J	1008	1/1	0.92	0.15	125,125,125,125	0
7	SO4	D	1101	5/5	0.95	0.09	51,51,54,55	0
6	MN	I	1003	1/1	0.96	0.24	83,83,83,83	0
6	MN	D	1007	1/1	0.97	0.34	93,93,93,93	0
7	SO4	C	1102	5/5	0.98	0.08	39,40,41,43	0
7	SO4	G	1103	5/5	0.98	0.12	42,43,44,46	0
6	MN	A	1001	1/1	0.99	0.18	35,35,35,35	0

6.5 Other polymers (i)

There are no such residues in this entry.