wwPDB X-ray Structure Validation Summary Report (i)

Jan 31, 2016 – 10:33 PM GMT

PDB ID	:	1U35
Title	:	Crystal structure of the nucleosome core particle containing the histone domain of macroH2A
Authors	:	Chakravarthy, S.; Gundimella, S.K.; Caron, C.; Perche, P.Y.; Pehrson, J.R.; Khochbin, S.; Luger, K.
Deposited on	:	2004-07-20
Resolution	:	3.00 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:

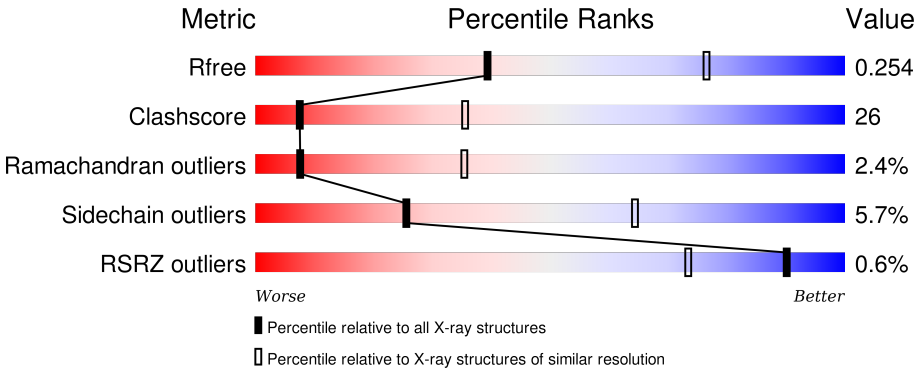
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 3.00 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.



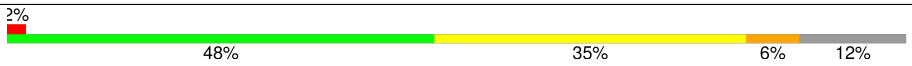
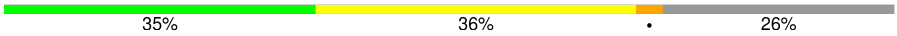
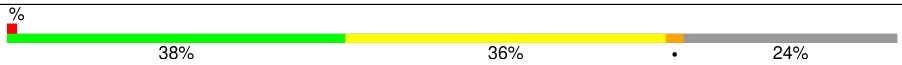
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	1578 (3.00-3.00)
Clashscore	102246	1912 (3.00-3.00)
Ramachandran outliers	100387	1853 (3.00-3.00)
Sidechain outliers	100360	1856 (3.00-3.00)
RSRZ outliers	91569	1592 (3.00-3.00)

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	I	146	32% 66%
1	J	146	38% 62%
2	A	136	36% 34% 28%
2	E	136	% 47% 23% 28%
3	B	103	44% 31% 23%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	103	
4	C	120	
4	G	120	
5	D	126	
5	H	126	

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 12053 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 DNA chain called alpha-satellite DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	I	145	Total	C	N	O	P	0	0	0
			2970	1420	539	867	144			
1	J	145	Total	C	N	O	P	0	0	0
			2969	1420	536	869	144			

- Molecule 2 is a protein called Histone H3.1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	A	98	Total	C	N	O	S	0	0	0
			807	508	156	139	4			
2	E	98	Total	C	N	O	S	0	0	0
			807	508	156	139	4			

- Molecule 3 is a protein called Hist1h4i protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	B	79	Total	C	N	O	S	0	0	0
			627	395	121	110	1			
3	F	83	Total	C	N	O	S	0	0	0
			662	418	129	114	1			

- Molecule 4 is a protein called H2A histone family.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	C	106	Total	C	N	O	S	0	0	0
			810	520	149	139	2			
4	G	106	Total	C	N	O	S	0	0	0
			810	520	149	139	2			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	867	VAL	GLY	ENGINEERED	UNP O75367
G	1067	VAL	GLY	ENGINEERED	UNP O75367

- Molecule 5 is a protein called histone 3, H2ba.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	D	93	Total	C	N	O	S	0	0	0
			731	459	131	139	2			
5	H	96	Total	C	N	O	S	0	0	0
			755	473	138	142	2			

- Molecule 6 is water.

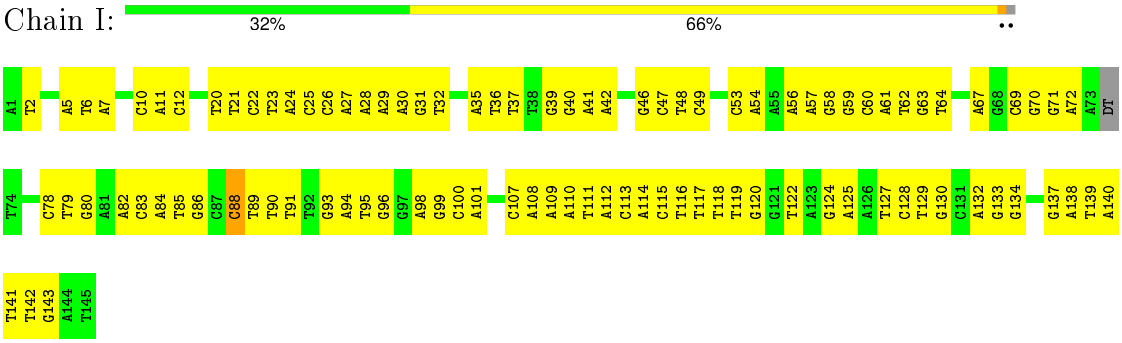
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	8	Total	O	0	0
			8	8		
6	B	6	Total	O	0	0
			6	6		
6	C	10	Total	O	0	0
			10	10		
6	D	9	Total	O	0	0
			9	9		
6	E	11	Total	O	0	0
			11	11		
6	F	13	Total	O	0	0
			13	13		
6	G	4	Total	O	0	0
			4	4		
6	H	4	Total	O	0	0
			4	4		
6	I	20	Total	O	0	0
			20	20		
6	J	20	Total	O	0	0
			20	20		

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.

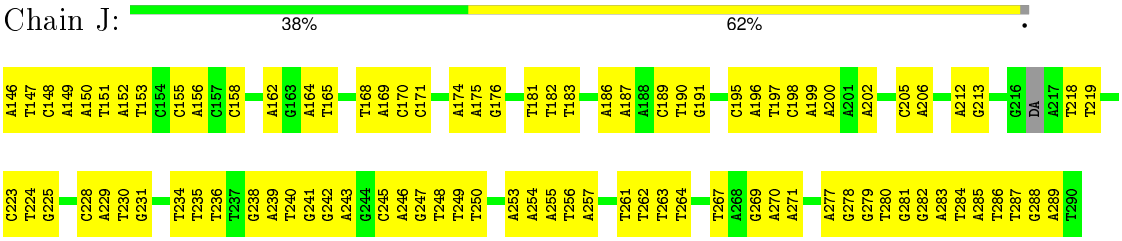
- Molecule 1: alpha-satellite DNA

Chain I:



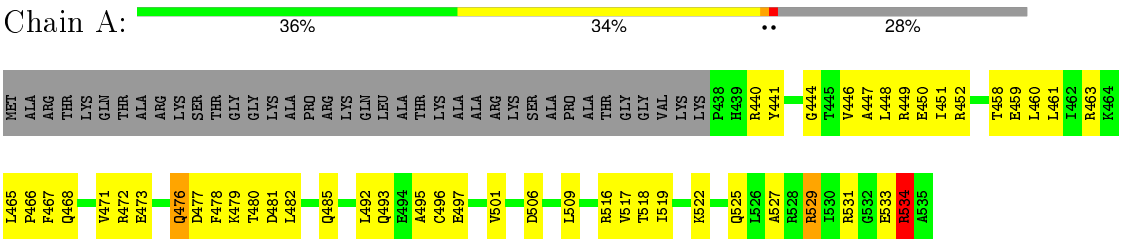
- Molecule 1: alpha-satellite DNA

Chain J:



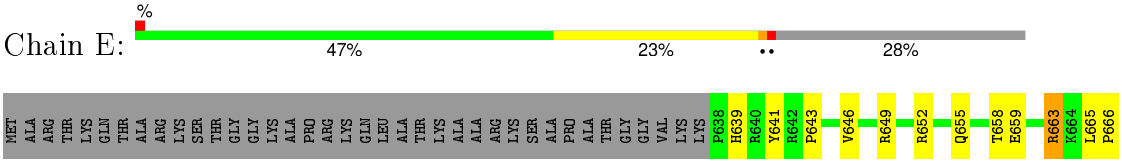
- Molecule 2: Histone H3.1

Chain A:



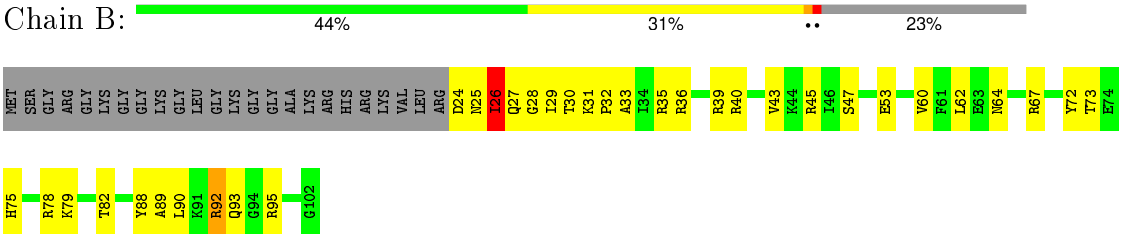
- Molecule 2: Histone H3.1

Chain E:



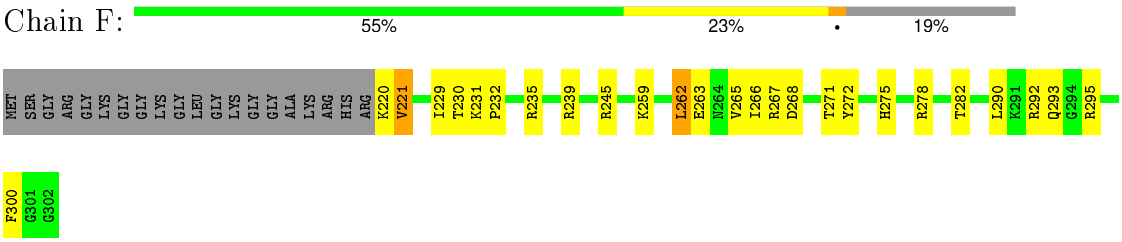
• Molecule 3: Hist1h4i protein

Chain B:



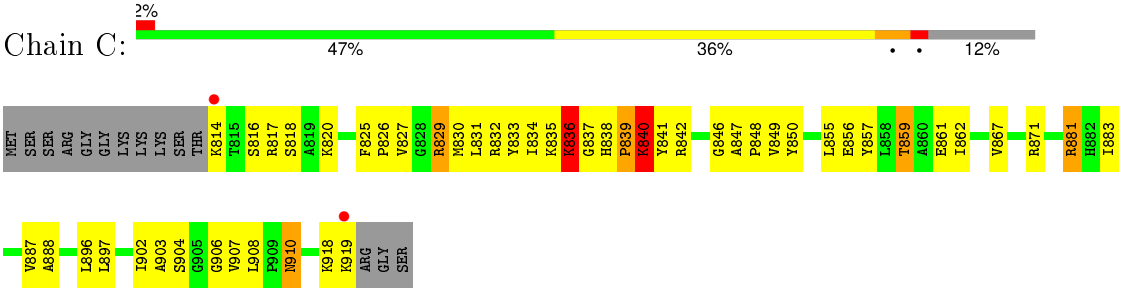
• Molecule 3: Hist1h4i protein

Chain F:



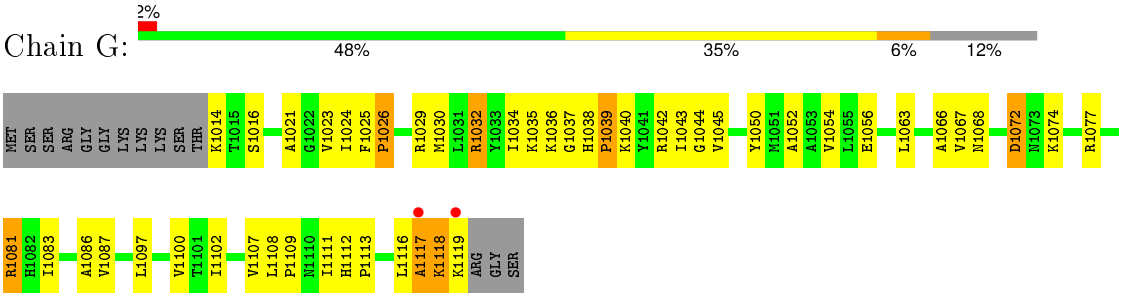
• Molecule 4: H2A histone family

Chain C:



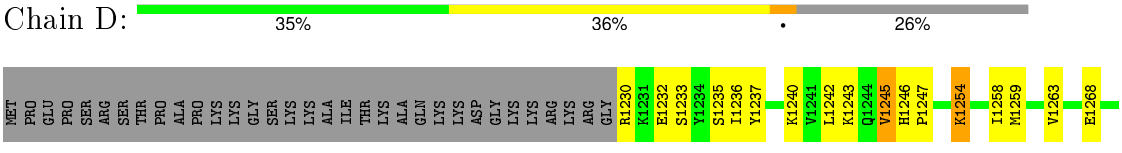
• Molecule 4: H2A histone family

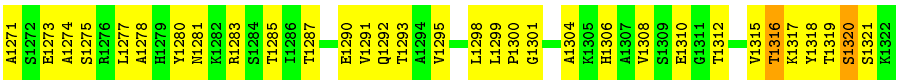
Chain G:



• Molecule 5: histone 3, H2ba

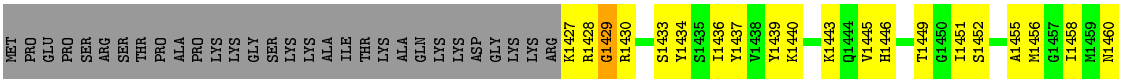
Chain D:

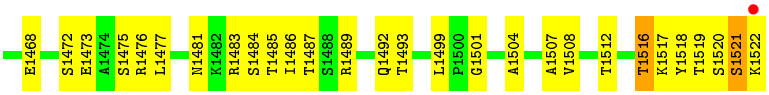




- Molecule 5: histone 3, H2ba

Chain H:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	105.50Å 109.60Å 175.99Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 3.00 48.63 – 2.95	Depositor EDS
% Data completeness (in resolution range)	95.7 (50.00-3.00) 95.1 (48.63-2.95)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.43 (at 2.96Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.206 , 0.260 0.204 , 0.254	Depositor DCC
R_{free} test set	2004 reflections (5.04%)	DCC
Wilson B-factor (Å^2)	78.6	Xtriage
Anisotropy	0.412	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 76.9	EDS
Estimated twinning fraction	0.018 for k,h,-l	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Outliers	0 of 43322 reflections	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	12053	wwPDB-VP
Average B, all atoms (Å^2)	97.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.79% 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

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	I	0.35	0/3332	0.69	0/5141
1	J	0.34	0/3330	0.69	0/5138
2	A	0.40	0/819	0.62	0/1097
2	E	0.48	0/819	0.71	1/1097 (0.1%)
3	B	0.43	0/634	0.77	1/848 (0.1%)
3	F	0.47	0/669	0.68	0/894
4	C	0.55	1/823 (0.1%)	0.74	0/1111
4	G	0.43	0/823	0.75	0/1111
5	D	0.45	0/742	0.68	0/996
5	H	0.40	0/766	0.60	0/1026
All	All	0.40	1/12757 (0.0%)	0.69	2/18459 (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
1	I	0	2
1	J	0	1
All	All	0	3

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	C	840	LYS	CE-NZ	7.45	1.67	1.49

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	B	26	ILE	CB-CA-C	-5.88	99.84	111.60
2	E	677	ASP	CB-CG-OD2	5.51	123.26	118.30

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	I	67	DA	Sidechain
1	I	88	DC	Sidechain
1	J	212	DA	Sidechain

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	I	2970	0	1639	132	0
1	J	2969	0	1640	106	0
2	A	807	0	844	58	0
2	E	807	0	844	43	1
3	B	627	0	663	51	0
3	F	662	0	709	30	0
4	C	810	0	873	94	0
4	G	810	0	873	87	0
5	D	731	0	753	51	0
5	H	755	0	782	50	0
6	A	8	0	0	1	0
6	B	6	0	0	1	0
6	C	10	0	0	2	0
6	D	9	0	0	1	1
6	E	11	0	0	2	0
6	F	13	0	0	1	0
6	G	4	0	0	1	0
6	H	4	0	0	1	0
6	I	20	0	0	1	0
6	J	20	0	0	2	0
All	All	12053	0	9620	547	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 26.

The worst 5 of 547 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)
4:C:840:LYS:CE	4:C:840:LYS:NZ	1.67	1.57
3:B:26:ILE:CG2	3:B:27:GLN:H	1.25	1.41
4:C:840:LYS:HG3	4:G:1039:PRO:CB	1.64	1.25
3:B:26:ILE:HG23	3:B:27:GLN:N	1.34	1.18
4:C:835:LYS:O	4:C:839:PRO:HD3	1.43	1.15

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)
2:E:677:ASP:OD2	6:D:301:HOH:O[3_745]	1.96	0.24

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
2	A	96/136 (71%)	92 (96%)	1 (1%)	3 (3%)	5 28
2	E	96/136 (71%)	92 (96%)	3 (3%)	1 (1%)	19 61
3	B	77/103 (75%)	72 (94%)	4 (5%)	1 (1%)	15 53
3	F	81/103 (79%)	78 (96%)	2 (2%)	1 (1%)	16 56
4	C	104/120 (87%)	93 (89%)	8 (8%)	3 (3%)	6 29
4	G	104/120 (87%)	94 (90%)	7 (7%)	3 (3%)	6 29
5	D	91/126 (72%)	79 (87%)	10 (11%)	2 (2%)	8 38
5	H	94/126 (75%)	79 (84%)	11 (12%)	4 (4%)	3 19
All	All	743/970 (77%)	679 (91%)	46 (6%)	18 (2%)	7 35

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	A	534	ARG

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	B	26	ILE
4	C	840	LYS
5	D	1320	SER
2	E	734	ARG

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	
2	A	85/111 (77%)	80 (94%)	5 (6%)	24	63
2	E	85/111 (77%)	83 (98%)	2 (2%)	57	87
3	B	64/79 (81%)	62 (97%)	2 (3%)	47	83
3	F	68/79 (86%)	65 (96%)	3 (4%)	35	74
4	C	84/95 (88%)	79 (94%)	5 (6%)	24	62
4	G	84/95 (88%)	78 (93%)	6 (7%)	18	54
5	D	81/108 (75%)	74 (91%)	7 (9%)	13	44
5	H	83/108 (77%)	77 (93%)	6 (7%)	18	53
All	All	634/786 (81%)	598 (94%)	36 (6%)	25	64

5 of 36 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
5	D	1298	LEU
2	E	734	ARG
5	H	1487	THR
5	D	1320	SER
3	F	245	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
2	E	639	HIS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	E	655	GLN
5	H	1460	ASN
5	D	1292	GLN
4	G	1068	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 carbohydrates 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	I	145/146 (99%)	-0.34	0 100 100	57, 128, 181, 200	0
1	J	145/146 (99%)	-0.38	0 100 100	60, 127, 168, 198	0
2	A	98/136 (72%)	-0.28	0 100 100	39, 66, 124, 151	0
2	E	98/136 (72%)	-0.32	1 (1%) 84 60	36, 52, 80, 169	0
3	B	79/103 (76%)	-0.28	0 100 100	46, 68, 104, 156	0
3	F	83/103 (80%)	-0.20	0 100 100	34, 51, 75, 151	0
4	C	106/120 (88%)	-0.19	2 (1%) 70 41	38, 60, 124, 195	0
4	G	106/120 (88%)	-0.25	2 (1%) 70 41	50, 70, 119, 162	0
5	D	93/126 (73%)	-0.28	0 100 100	44, 64, 104, 150	0
5	H	96/126 (76%)	-0.23	1 (1%) 84 60	43, 72, 131, 159	0
All	All	1049/1262 (83%)	-0.28	6 (0%) 90 73	34, 71, 156, 200	0

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	C	919	LYS	4.9
4	C	814	LYS	4.5
4	G	1117	ALA	3.2
5	H	1522	LYS	3.1
2	E	735	ALA	2.9

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 carbohydrates 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.