



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 28, 2023 – 05:57 AM EDT

PDB ID : 3KXB  
Title : Structural characterization of H3K56Q nucleosomes and nucleosomal arrays  
Authors : Clark, N.J.; Lilyestrom, W.G.  
Deposited on : 2009-12-02  
Resolution : 3.20 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](mailto: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 : 4.02b-467  
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

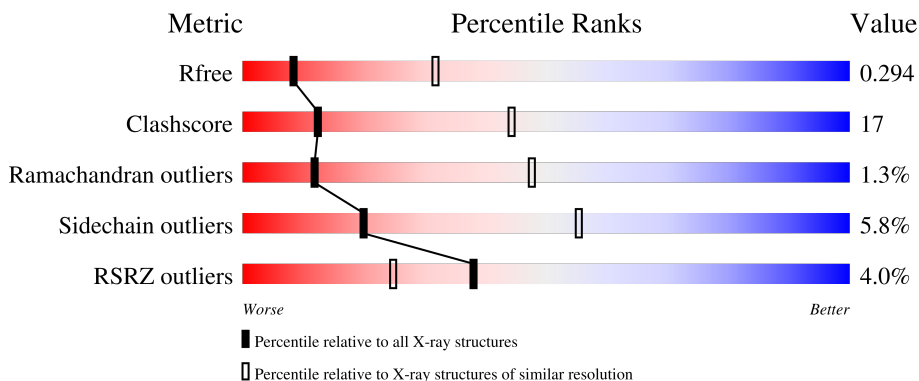
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.20 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 of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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	135	 2% 44% 19% 5% 32%
1	E	135	 43% 23% 5% 33%
2	B	102	 57% 14% 5% 27%
2	F	102	 2% 55% 22% 5% 16%
3	C	129	 59% 19% 5% 19%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	G	129	<p>2% 58% 16% 22%</p>
4	D	122	<p>2% 53% 16% 30%</p>
4	H	122	<p>2% 58% 15% 22%</p>
5	I	146	<p>15% 55% 45%</p>
5	J	146	<p>5% 55% 45%</p>

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 11868 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
			Total	C	N	O	S			
1	A	92	751	473	141	134	3	0	0	0
1	E	90	740	466	139	132	3	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	56	GLU	LYS	engineered mutation	UNP P84233
A	102	ALA	GLY	engineered mutation	UNP P84233
E	56	GLU	LYS	engineered mutation	UNP P84233
E	102	ALA	GLY	engineered mutation	UNP P84233

- Molecule 2 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	74	590	374	114	101	1	0	0	0
2	F	86	694	436	140	117	1	0	0	0

- Molecule 3 is a protein called Histone H2A.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
3	C	105	809	510	158	141	0	0	0
3	G	100	774	491	150	133	0	0	0

- Molecule 4 is a protein called Histone H2B 1.1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	86	Total	C	N	O	S	0	0	0
			669	421	120	126	2			
4	H	95	Total	C	N	O	S	0	0	0
			746	469	136	139	2			

There are 2 discrepancies between the modelled and reference sequences:

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

- Molecule 5 is a DNA chain called PALINDROMIC 146 BP DNA REPEAT 8/9 FROM HUMAN X-CHROMOSOME ALPHA SATELLITE DNA.

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			

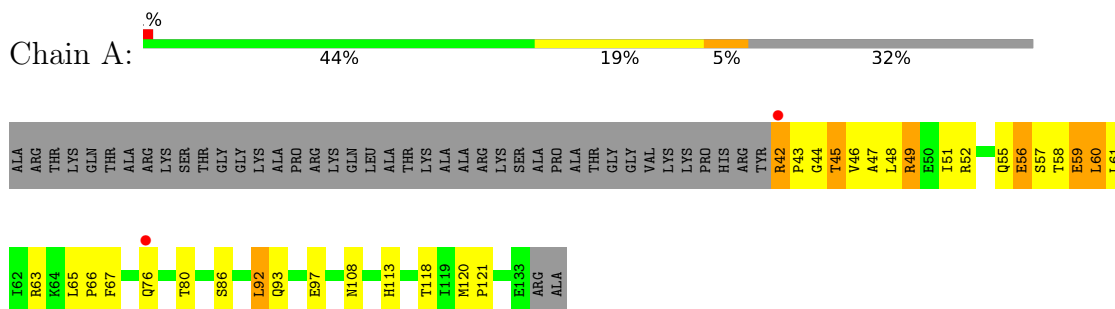
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	9	Total	O	0	0
			9	9		
6	B	6	Total	O	0	0
			6	6		
6	C	10	Total	O	0	0
			10	10		
6	D	4	Total	O	0	0
			4	4		
6	E	9	Total	O	0	0
			9	9		
6	F	7	Total	O	0	0
			7	7		
6	G	10	Total	O	0	0
			10	10		
6	H	4	Total	O	0	0
			4	4		
6	I	24	Total	O	0	0
			24	24		
6	J	32	Total	O	0	0
			32	32		

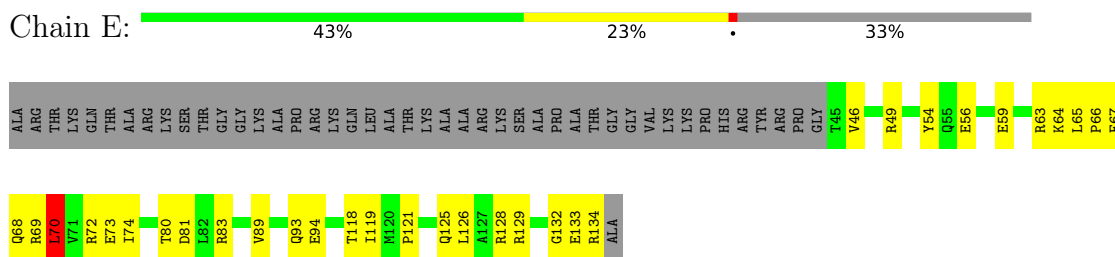
### 3 Residue-property plots [i](#)

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.

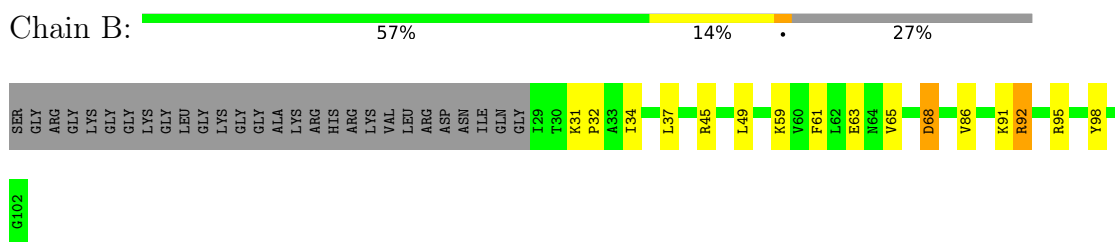
- Molecule 1: Histone H3.2



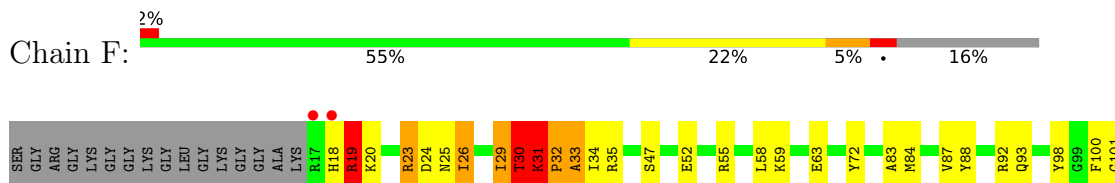
- Molecule 1: Histone H3.2



- Molecule 2: Histone H4

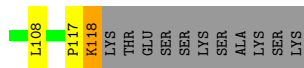
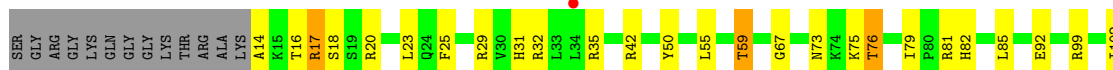


- Molecule 2: Histone H4

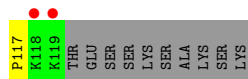
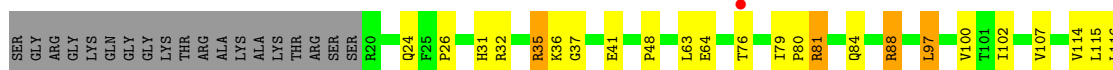


G102

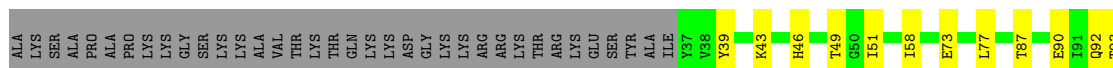
- Molecule 3: Histone H2A



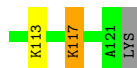
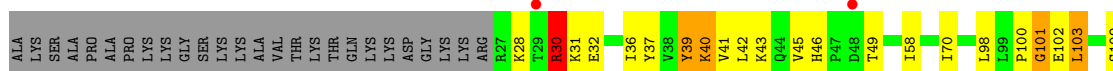
- Molecule 3: Histone H2A



- Molecule 4: Histone H2B 1.1

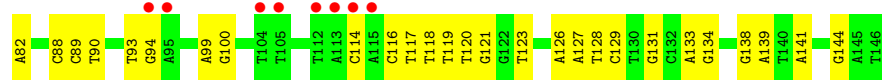
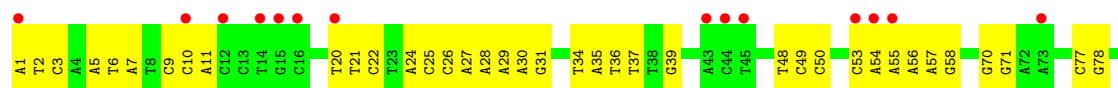


- Molecule 4: Histone H2B 1.1

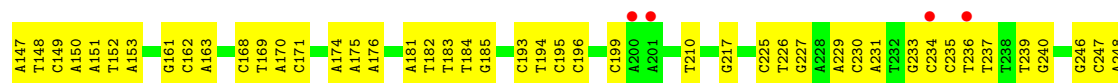


- Molecule 5: PALINDROMIC 146 BP DNA REPEAT 8/9 FROM HUMAN X-CHROMOSOME ALPHA SATELLITE DNA





● Molecule 5: PALINDROMIC 146 BP DNA REPEAT 8/9 FROM HUMAN X-CHROMOSOME ALPHA SATELLITE DNA





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	109.52Å 105.67Å 181.33Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 3.20 33.53 – 3.20	Depositor EDS
% Data completeness (in resolution range)	78.0 (20.00-3.20) 77.8 (33.53-3.20)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	0.06	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.67 (at 3.18Å)	Xtrriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.283 , 0.292 0.284 , 0.294	Depositor DCC
$R_{free}$ test set	2893 reflections (9.96%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	64.6	Xtrriage
Anisotropy	0.270	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.25 , 53.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.28$	Xtrriage
Estimated twinning fraction	0.034 for k,h,-l	Xtrriage
$F_o, F_c$ correlation	0.86	EDS
Total number of atoms	11868	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	86.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.14% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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 [i](#)

### 5.1 Standard geometry [i](#)

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	0.33	0/760	0.52	0/1021
1	E	0.37	0/748	0.58	0/1004
2	B	0.42	1/597 (0.2%)	0.56	0/798
2	F	0.36	0/702	0.48	0/937
3	C	0.37	0/819	0.54	0/1106
3	G	0.43	1/784 (0.1%)	0.59	1/1059 (0.1%)
4	D	0.39	0/679	0.56	0/913
4	H	0.38	0/757	0.49	0/1018
5	I	0.31	0/3354	0.69	0/5175
5	J	0.31	0/3354	0.69	0/5175
All	All	0.35	2/12554 (0.0%)	0.63	1/18206 (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	E	0	2
2	F	0	5
3	C	0	1
3	G	0	1
4	D	0	1
4	H	0	1
All	All	0	11

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	G	97	LEU	C-N	7.92	1.47	1.33
2	B	68	ASP	C-N	5.01	1.45	1.34

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	G	97	LEU	O-C-N	-5.88	113.20	123.20

There are no chirality outliers.

5 of 11 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	C	14	ALA	Peptide
4	D	121	ALA	Peptide
1	E	132	GLY	Peptide
1	E	70	LEU	Peptide
2	F	29	ILE	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	751	0	784	35	0
1	E	740	0	774	44	0
2	B	590	0	631	18	0
2	F	694	0	742	54	0
3	C	809	0	864	26	0
3	G	774	0	827	32	0
4	D	669	0	691	22	4
4	H	746	0	773	33	0
5	I	2990	0	1651	75	0
5	J	2990	0	1651	77	4
6	A	9	0	0	0	0
6	B	6	0	0	0	0
6	C	10	0	0	2	0
6	D	4	0	0	0	0
6	E	9	0	0	0	0
6	F	7	0	0	1	0
6	G	10	0	0	1	0
6	H	4	0	0	1	0
6	I	24	0	0	9	0
6	J	32	0	0	18	0
All	All	11868	0	9388	357	4

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 17.

The worst 5 of 357 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)
3:G:35:ARG:NH2	3:G:41:GLU:O	1.57	1.38
2:F:72:TYR:HB3	6:F:106:HOH:O	1.20	1.28
2:B:92:ARG:NH2	4:D:97:LEU:O	1.66	1.27
3:G:24:GLN:O	4:H:37:TYR:CD1	1.86	1.27
1:E:68:GLN:CG	1:E:89:VAL:HG11	1.74	1.15

All (4) 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)
4:D:122:LYS:N	5:J:210:DT:OP1[4_465]	1.88	0.32
4:D:122:LYS:CD	5:J:210:DT:C4'[4_465]	1.96	0.24
4:D:122:LYS:CB	5:J:210:DT:C5'[4_465]	2.10	0.10
4:D:122:LYS:NZ	5:J:210:DT:O3'[4_465]	2.15	0.05

## 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	90/135 (67%)	82 (91%)	7 (8%)	1 (1%)	14	51
1	E	88/135 (65%)	83 (94%)	5 (6%)	0	100	100
2	B	72/102 (71%)	71 (99%)	1 (1%)	0	100	100
2	F	84/102 (82%)	74 (88%)	6 (7%)	4 (5%)	2	17
3	C	103/129 (80%)	99 (96%)	4 (4%)	0	100	100
3	G	98/129 (76%)	93 (95%)	5 (5%)	0	100	100

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
4	D	84/122 (69%)	82 (98%)	1 (1%)	1 (1%)	13	49
4	H	93/122 (76%)	87 (94%)	3 (3%)	3 (3%)	4	26
All	All	712/976 (73%)	671 (94%)	32 (4%)	9 (1%)	12	47

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	F	31	LYS
4	D	101	GLY
2	F	19	ARG
2	F	32	PRO
4	H	39	TYR

### 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	80/110 (73%)	73 (91%)	7 (9%)	10	36
1	E	79/110 (72%)	78 (99%)	1 (1%)	69	87
2	B	60/78 (77%)	56 (93%)	4 (7%)	16	50
2	F	71/78 (91%)	63 (89%)	8 (11%)	6	25
3	C	83/101 (82%)	76 (92%)	7 (8%)	11	39
3	G	79/101 (78%)	76 (96%)	3 (4%)	33	67
4	D	73/102 (72%)	73 (100%)	0	100	100
4	H	81/102 (79%)	76 (94%)	5 (6%)	18	53
All	All	606/782 (78%)	571 (94%)	35 (6%)	20	55

5 of 35 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	G	81	ARG
3	G	88	ARG

Continued on next page...

*Continued from previous page...*

Mol	Chain	Res	Type
4	H	102	GLU
3	C	18	SER
3	C	17	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
4	D	92	GLN
4	H	46	HIS
4	H	79	HIS
1	A	108	ASN
1	A	76	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 monosaccharides 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	92/135 (68%)	-0.02	2 (2%) 62 48	33, 61, 78, 107	0
1	E	90/135 (66%)	-0.09	0 100 100	30, 45, 61, 78	0
2	B	74/102 (72%)	-0.29	0 100 100	42, 54, 69, 76	0
2	F	86/102 (84%)	-0.19	2 (2%) 60 47	28, 45, 67, 108	0
3	C	105/129 (81%)	-0.13	1 (0%) 82 72	28, 47, 82, 91	0
3	G	100/129 (77%)	-0.08	3 (3%) 50 34	44, 61, 86, 103	0
4	D	86/122 (70%)	-0.16	2 (2%) 60 47	28, 52, 71, 95	0
4	H	95/122 (77%)	0.09	2 (2%) 63 49	41, 69, 92, 127	0
5	I	146/146 (100%)	0.87	22 (15%) 2 1	69, 115, 155, 180	0
5	J	146/146 (100%)	0.55	7 (4%) 30 18	68, 114, 156, 162	0
All	All	1020/1268 (80%)	0.13	41 (4%) 38 25	28, 63, 142, 180	0

The worst 5 of 41 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	I	14	DT	5.5
5	I	15	DG	5.0
5	I	45	DT	4.0
5	I	55	DA	4.0
5	I	44	DC	3.7

### 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 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.