



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 17, 2024 – 01:09 AM EDT

PDB ID : 5OXV  
Title : Structure of the 4\_601\_157 tetranucleosome (C2 form)  
Authors : Ekundayo, B.; Schalch, T.  
Deposited on : 2017-09-07  
Resolution : 6.72 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	:	2.37.1
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)	:	2.37.1

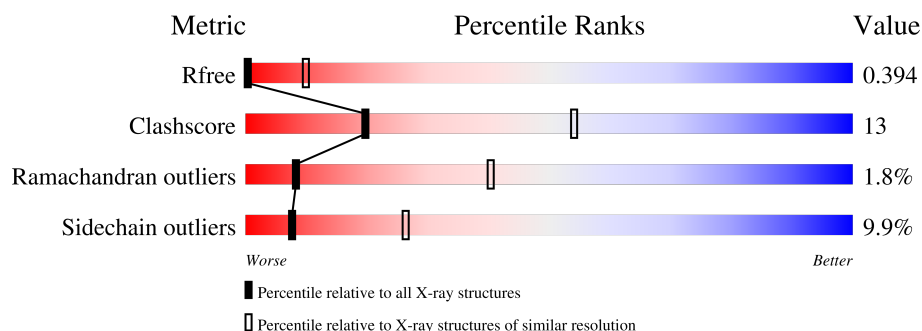
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 6.72 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	1002 (9.50-3.90)
Clashscore	141614	1066 (9.50-3.90)
Ramachandran outliers	138981	1000 (9.50-3.90)
Sidechain outliers	138945	1000 (9.50-3.86)

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  $\leq 5\%$ .

Mol	Chain	Length	Quality of chain
1	J	312	44% 44% 12% .
2	D	126	43% 27% 5% . 25%
2	H	126	44% 27% . 26%
2	N	126	44% 26% 5% . 25%
2	R	126	44% 26% . 26%
3	C	130	51% 24% 5% 21%
3	G	130	48% 28% 5% 19%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	M	130	
3	Q	130	
4	B	102	
4	F	102	
4	L	102	
4	P	102	
5	A	135	
5	E	135	
5	K	135	
5	O	135	
6	I	313	

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 24648 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 DNA STRAND 1 (601-based sequence model).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	J	310	Total	C	N	O	P	0	0	0
			6394	3024	1203	1857	310			

- Molecule 2 is a protein called Histone H2B 1.1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	R	93	Total	C	N	O	S	0	0	0
			726	457	130	137	2			
2	N	95	Total	C	N	O	S	0	0	0
			745	469	134	140	2			
2	H	93	Total	C	N	O	S	0	0	0
			726	457	130	137	2			
2	D	95	Total	C	N	O	S	0	0	0
			745	469	134	140	2			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
R	29	THR	SER	conflict	UNP P02281
N	29	THR	SER	conflict	UNP P02281
H	29	THR	SER	conflict	UNP P02281
D	29	THR	SER	conflict	UNP P02281

- Molecule 3 is a protein called Histone H2A.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	Q	105	Total	C	N	O	0	0	0
			809	510	158	141			
3	M	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			

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	103	Total	C	N	O	0	0	0
			795	501	155	139			

- Molecule 4 is a protein called Histone H4.

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

- Molecule 5 is a protein called Histone H3.2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	O	97	Total	C	N	O	S	0	0	0
			801	504	155	139	3			
5	K	97	Total	C	N	O	S	0	0	0
			802	506	155	138	3			
5	E	97	Total	C	N	O	S	0	0	0
			801	504	155	139	3			
5	A	97	Total	C	N	O	S	0	0	0
			802	506	155	138	3			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
O	102	ALA	GLY	conflict	UNP P84233
K	102	ALA	GLY	conflict	UNP P84233
E	102	ALA	GLY	conflict	UNP P84233
A	102	ALA	GLY	conflict	UNP P84233

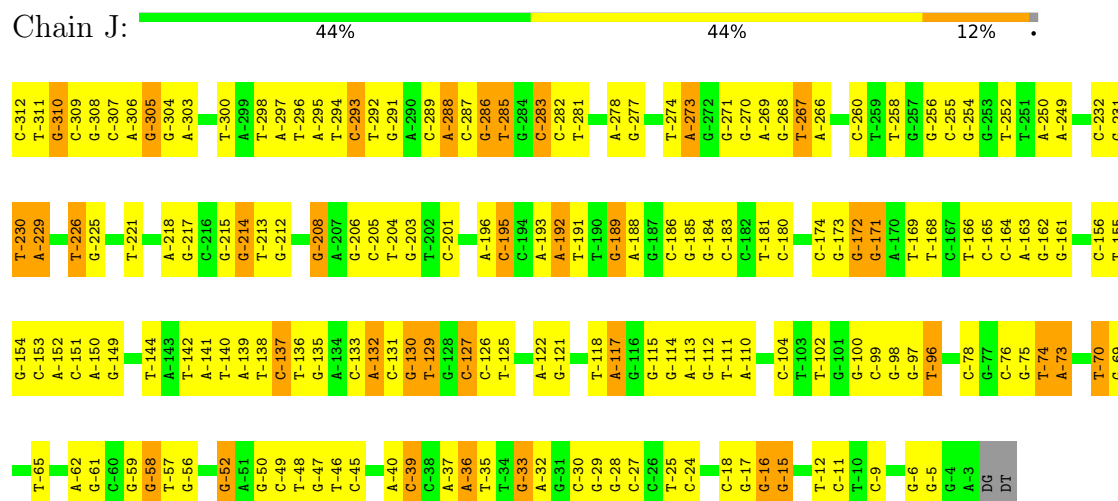
- Molecule 6 is a DNA chain called DNA STRAND 2 (601-based sequence model).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	I	311	Total	C	N	O	P	0	0	0
			6336	3009	1146	1870	311			

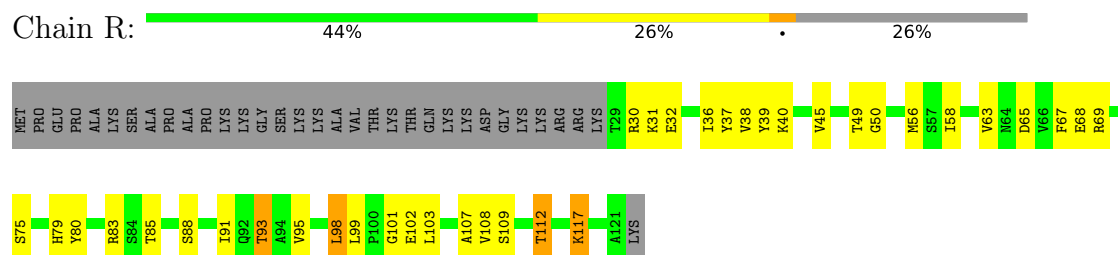
### 3 Residue-property plots [i](#)

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. 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. 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.

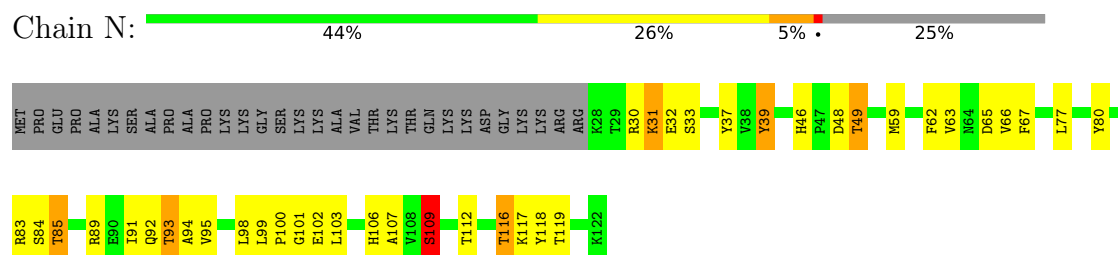
- Molecule 1: DNA STRAND 1 (601-based sequence model)



- Molecule 2: Histone H2B 1.1



- Molecule 2: Histone H2B 1.1



- Molecule 2: Histone H2B 1.1

- Molecule 2: Histone H2B 1.1

- Molecule 3: Histone H2A

- Molecule 3: Histone H2A

- Molecule 3: Histone H2A

- Molecule 3: Histone H2A

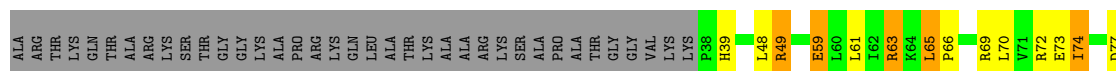






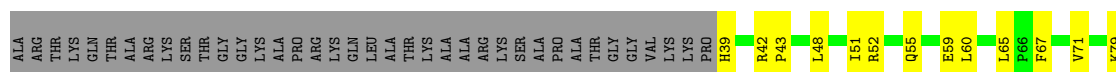
- Molecule 5: Histone H3.2

Chain K: 49% 19% 28%



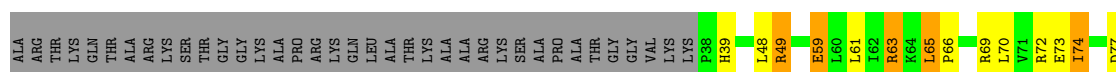
- Molecule 5: Histone H3.2

Chain E: 52% 20% 28%



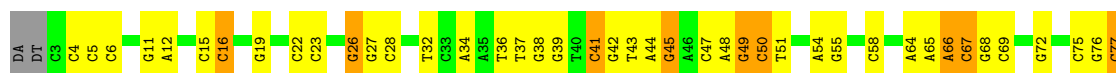
- Molecule 5: Histone H3.2

Chain A: 49% 19% 28%



- Molecule 6: DNA STRAND 2 (601-based sequence model)

Chain I: 45% 42% 12%



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	348.84Å 63.27Å 271.36Å 90.00° 124.74° 90.00°	Depositor
Resolution (Å)	71.67 – 6.72 222.99 – 6.72	Depositor EDS
% Data completeness (in resolution range)	87.8 (71.67-6.72) 82.3 (222.99-6.72)	Depositor EDS
$R_{merge}$	0.16	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.22 (at 6.74Å)	Xtriage
Refinement program	PHENIX (1.11.1 _2575: ???)	Depositor
R, $R_{free}$	0.317 , 0.352 0.358 , 0.394	Depositor DCC
$R_{free}$ test set	1169 reflections (12.72%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	308.9	Xtriage
Anisotropy	0.465	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	1.04 , -10.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.45$ , $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.79	EDS
Total number of atoms	24648	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	637.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.99% 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	J	0.74	2/7182 (0.0%)	1.47	96/11093 (0.9%)
2	D	0.49	0/756	0.64	0/1015
2	H	0.62	0/737	0.70	0/993
2	N	0.49	0/756	0.64	0/1015
2	R	0.62	0/737	0.70	0/993
3	C	0.42	0/805	0.61	0/1088
3	G	0.62	0/819	0.76	0/1106
3	M	0.43	0/805	0.61	0/1088
3	Q	0.63	0/819	0.76	0/1106
4	B	0.67	0/669	0.84	0/894
4	F	0.46	0/626	0.61	0/837
4	L	0.67	0/669	0.84	0/894
4	P	0.45	0/626	0.61	0/837
5	A	0.66	0/814	0.73	0/1092
5	E	0.41	0/812	0.58	0/1088
5	K	0.66	0/814	0.73	0/1092
5	O	0.41	0/812	0.58	0/1088
6	I	0.74	4/7098 (0.1%)	1.38	84/10943 (0.8%)
All	All	0.66	6/26356 (0.0%)	1.17	180/38262 (0.5%)

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	J	-310	DG	C1'-N9	-6.27	1.38	1.47
6	I	310	DC	C1'-N1	5.73	1.56	1.49
1	J	-305	DG	C1'-N9	-5.53	1.39	1.47
6	I	312	DG	C1'-N9	-5.43	1.39	1.47
6	I	308	DT	C1'-N1	5.30	1.56	1.49

The worst 5 of 180 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	J	-163	DA	O3'-P-O5'	33.99	168.59	104.00

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	J	-163	DA	P-O3'-C3'	29.81	155.47	119.70
1	J	-163	DA	OP1-P-O3'	-20.17	60.83	105.20
6	I	162	DC	O3'-P-O5'	14.02	130.63	104.00
1	J	-163	DA	OP2-P-O3'	-12.28	78.18	105.20

There are no chirality outliers.

There are no planarity outliers.

## 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	J	6394	0	3474	141	2
2	D	745	0	773	37	3
2	H	726	0	747	36	2
2	N	745	0	773	36	2
2	R	726	0	747	35	2
3	C	795	0	846	25	0
3	G	809	0	864	43	1
3	M	795	0	846	26	1
3	Q	809	0	864	44	1
4	B	662	0	709	36	0
4	F	619	0	659	16	0
4	L	662	0	709	36	0
4	P	619	0	659	17	0
5	A	802	0	841	22	0
5	E	801	0	838	15	0
5	K	802	0	841	21	0
5	O	801	0	838	17	0
6	I	6336	0	3489	146	2
All	All	24648	0	19517	554	8

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 13.

The worst 5 of 554 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:N:93:THR:HG21	4:L:75:HIS:CD2	1.58	1.37
2:D:93:THR:HG21	4:B:75:HIS:CD2	1.58	1.37
1:J:-306:DA:C2	6:I:307:DG:N2	1.97	1.32
1:J:-306:DA:H2	6:I:307:DG:C2	1.48	1.29
1:J:-306:DA:C2	6:I:307:DG:C2	2.28	1.21

The worst 5 of 8 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:H:117:LYS:CE	2:D:117:LYS:NZ[1_545]	1.23	0.97
1:J:-312:DC:P	6:I:313:DT:O3'[2_656]	1.61	0.59
1:J:-312:DC:OP2	6:I:313:DT:O3'[2_656]	1.79	0.41
2:H:117:LYS:CE	2:D:117:LYS:CE[1_545]	1.86	0.34
2:R:117:LYS:CE	2:N:117:LYS:NZ[1_565]	1.90	0.30

## 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	D	93/126 (74%)	82 (88%)	9 (10%)	2 (2%)	6	35
2	H	91/126 (72%)	79 (87%)	11 (12%)	1 (1%)	14	52
2	N	93/126 (74%)	81 (87%)	10 (11%)	2 (2%)	6	35
2	R	91/126 (72%)	80 (88%)	10 (11%)	1 (1%)	14	52
3	C	101/130 (78%)	88 (87%)	9 (9%)	4 (4%)	3	23
3	G	103/130 (79%)	88 (85%)	15 (15%)	0	100	100
3	M	101/130 (78%)	88 (87%)	9 (9%)	4 (4%)	3	23
3	Q	103/130 (79%)	88 (85%)	15 (15%)	0	100	100
4	B	81/102 (79%)	70 (86%)	8 (10%)	3 (4%)	3	24
4	F	76/102 (74%)	67 (88%)	9 (12%)	0	100	100

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
4	L	81/102 (79%)	70 (86%)	8 (10%)	3 (4%)	3	24
4	P	76/102 (74%)	67 (88%)	9 (12%)	0	100	100
5	A	95/135 (70%)	83 (87%)	9 (10%)	3 (3%)	4	26
5	E	95/135 (70%)	86 (90%)	9 (10%)	0	100	100
5	K	95/135 (70%)	83 (87%)	9 (10%)	3 (3%)	4	26
5	O	95/135 (70%)	86 (90%)	9 (10%)	0	100	100
All	All	1470/1972 (74%)	1286 (88%)	158 (11%)	26 (2%)	8	40

5 of 26 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	N	101	GLY
4	L	29	ILE
5	K	73	GLU
2	D	101	GLY
4	B	29	ILE

### 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	
2	D	81/106 (76%)	69 (85%)	12 (15%)	3	15
2	H	79/106 (74%)	67 (85%)	12 (15%)	3	14
2	N	81/106 (76%)	69 (85%)	12 (15%)	3	15
2	R	79/106 (74%)	67 (85%)	12 (15%)	3	14
3	C	82/102 (80%)	75 (92%)	7 (8%)	10	33
3	G	83/102 (81%)	74 (89%)	9 (11%)	6	23
3	M	82/102 (80%)	75 (92%)	7 (8%)	10	33
3	Q	83/102 (81%)	74 (89%)	9 (11%)	6	23
4	B	68/78 (87%)	63 (93%)	5 (7%)	13	38
4	F	63/78 (81%)	59 (94%)	4 (6%)	18	43

Continued on next page...

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
4	L	68/78 (87%)	63 (93%)	5 (7%)	13	38
4	P	63/78 (81%)	59 (94%)	4 (6%)	18	43
5	A	85/110 (77%)	77 (91%)	8 (9%)	8	28
5	E	84/110 (76%)	79 (94%)	5 (6%)	19	44
5	K	85/110 (77%)	77 (91%)	8 (9%)	8	28
5	O	84/110 (76%)	79 (94%)	5 (6%)	19	44
All	All	1250/1584 (79%)	1126 (90%)	124 (10%)	8	26

5 of 124 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
5	K	63	ARG
3	C	118	LYS
2	H	109	SER
3	C	114	VAL
5	A	59	GLU

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

Mol	Chain	Res	Type
3	G	84	GLN
5	E	68	GLN
5	A	39	HIS
3	C	112	GLN
4	B	75	HIS

### 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 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 ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.3 Carbohydrates ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.4 Ligands ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.5 Other polymers ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.