



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 04:40 AM GMT

PDB ID : 2NQB  
Title : Drosophila Nucleosome Structure  
Authors : Luger, K.; Chakravarthy, S.  
Deposited on : 2006-10-30  
Resolution : 2.30 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  
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) 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

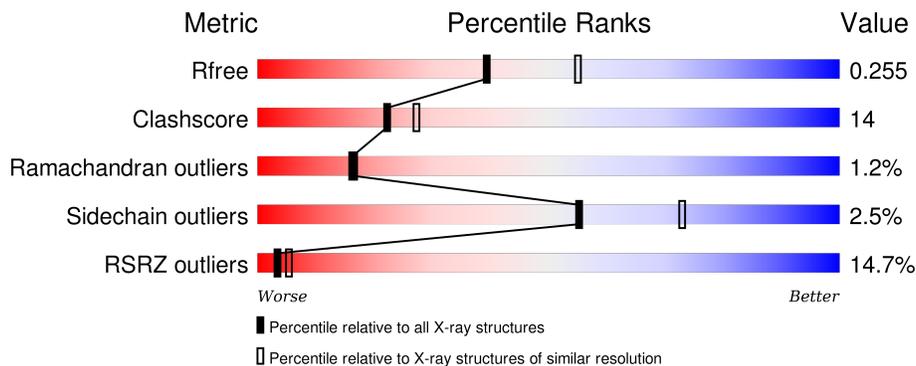
# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.30 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	3852 (2.30-2.30)
Clashscore	102246	4452 (2.30-2.30)
Ramachandran outliers	100387	4410 (2.30-2.30)
Sidechain outliers	100360	4409 (2.30-2.30)
RSRZ outliers	91569	3857 (2.30-2.30)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	I	146	
1	J	146	
2	A	135	
2	E	135	
3	B	103	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	F	103	<p>15% 58% 21% •• 17%</p>
4	C	123	<p>7% 63% 20% •• 14%</p>
4	G	123	<p>5% 63% 21% • 15%</p>
5	D	123	<p>10% 63% 13% • 23%</p>
5	H	123	<p>10% 63% 11% • 24%</p>

## 2 Entry composition i

There are 6 unique types of molecules in this entry. The entry contains 12294 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
			Total	C	N	O	P			
1	I	146	2990	1430	541	874	145	0	0	0
1	J	146	2990	1430	541	874	145	0	0	0

- Molecule 2 is a protein called Histone H3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	A	98	807	508	156	140	3	0	0	0
2	E	98	807	508	156	140	3	0	0	0

- Molecule 3 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	B	81	646	407	126	112	1	0	0	0
3	F	86	694	436	140	117	1	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	ILE	-	EXPRESSION TAG	UNP P84040
F	200	ILE	-	EXPRESSION TAG	UNP P84040

- Molecule 4 is a protein called Histone H2A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	C	106	812	512	158	141	1	0	0	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	G	105	Total	C	N	O	S	0	0	0
			803	506	156	140	1			

- Molecule 5 is a protein called Histone H2B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	D	95	Total	C	N	O	S	0	0	0
			751	472	136	141	2			
5	H	93	Total	C	N	O	S	0	0	0
			730	460	130	138	2			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	1200	ILE	-	EXPRESSION TAG	UNP P02283
D	1240	THR	LYS	ENGINEERED	UNP P02283
H	1400	ILE	-	EXPRESSION TAG	UNP P02283
H	1440	THR	LYS	ENGINEERED	UNP P02283

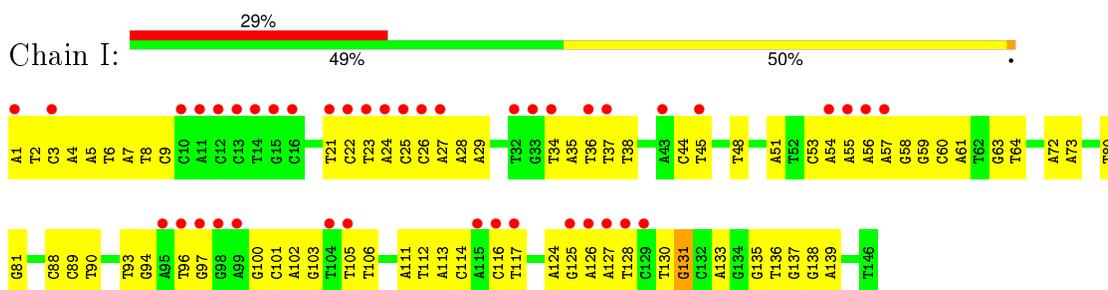
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	23	Total	O	0	0
			23	23		
6	B	16	Total	O	0	0
			16	16		
6	C	29	Total	O	0	0
			29	29		
6	D	28	Total	O	0	0
			28	28		
6	E	41	Total	O	0	0
			41	41		
6	F	26	Total	O	0	0
			26	26		
6	G	26	Total	O	0	0
			26	26		
6	H	11	Total	O	0	0
			11	11		
6	I	32	Total	O	0	0
			32	32		
6	J	32	Total	O	0	0
			32	32		

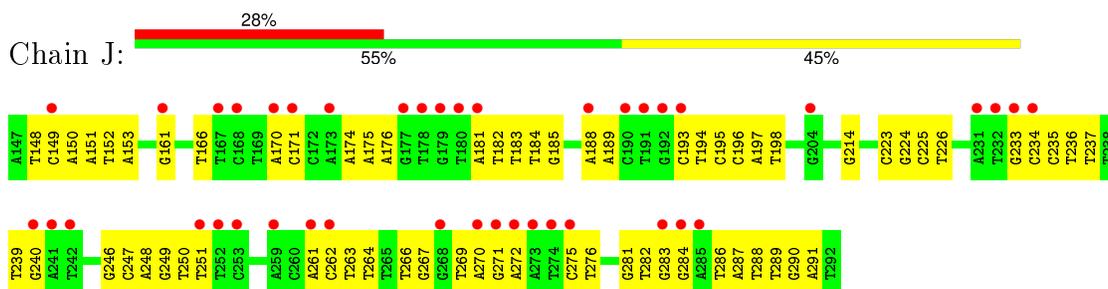
### 3 Residue-property plots

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.

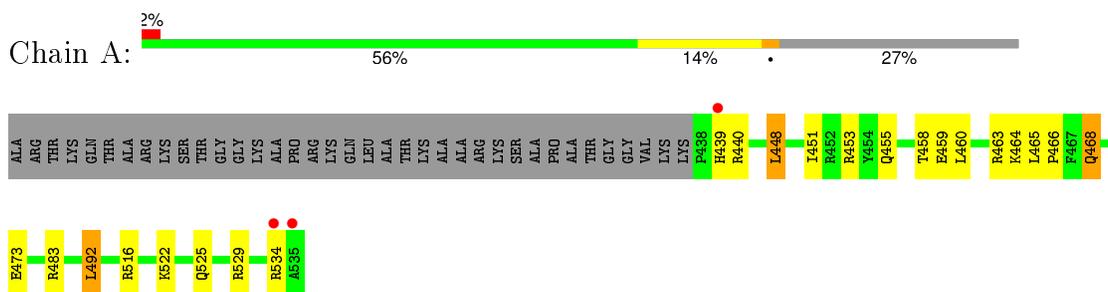
- Molecule 1: alpha-satellite DNA



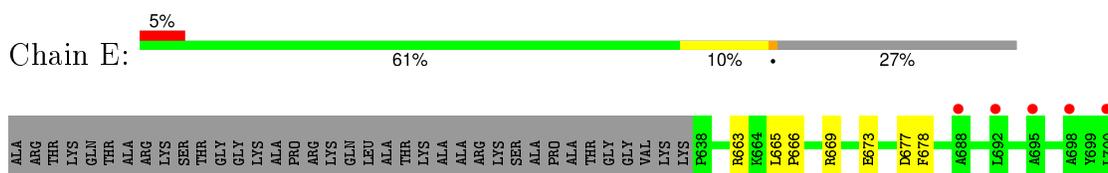
- Molecule 1: alpha-satellite DNA



- Molecule 2: Histone H3



- Molecule 2: Histone H3





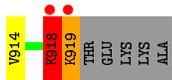
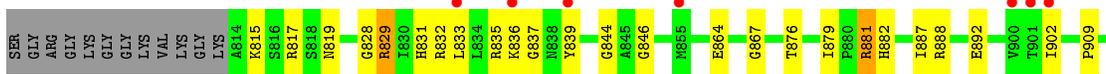
- Molecule 3: Histone H4



- Molecule 3: Histone H4



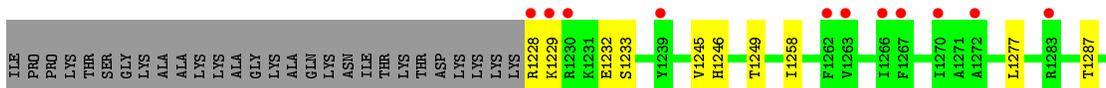
- Molecule 4: Histone H2A



- Molecule 4: Histone H2A

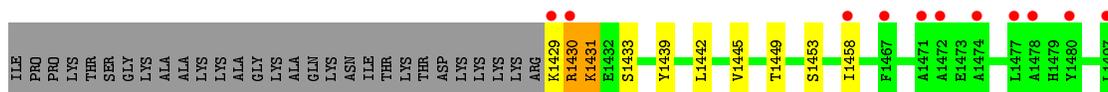


- Molecule 5: Histone H2B





● Molecule 5: Histone H2B



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	106.14Å 109.58Å 182.04Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	99.00 – 2.30 48.69 – 2.30	Depositor EDS
% Data completeness (in resolution range)	96.1 (99.00-2.30) 95.7 (48.69-2.30)	Depositor EDS
$R_{merge}$	0.04	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.33 (at 2.29Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.223 , 0.254 0.224 , 0.255	Depositor DCC
$R_{free}$ test set	2267 reflections (2.49%)	DCC
Wilson B-factor (Å <sup>2</sup> )	47.3	Xtriage
Anisotropy	0.267	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 65.7	EDS
Estimated twinning fraction	0.018 for k,h,l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 93893 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	12294	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	75.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.76% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	I	0.28	0/3354	0.68	0/5175
1	J	0.28	0/3354	0.67	0/5175
2	A	0.35	0/819	0.57	0/1097
2	E	0.44	0/819	0.59	0/1097
3	B	0.39	0/653	0.62	0/873
3	F	0.43	0/702	0.72	1/937 (0.1%)
4	C	0.38	0/822	0.56	0/1108
4	G	0.33	0/813	0.56	0/1097
5	D	0.41	0/762	0.57	0/1023
5	H	0.37	0/741	0.59	0/998
All	All	0.33	0/12839	0.64	1/18580 (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	1
1	J	0	1
All	All	0	2

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	F	220	LYS	N-CA-C	-5.07	97.31	111.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	I	131	DG	Sidechain
1	J	214	DG	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	2990	0	1651	87	0
1	J	2990	0	1651	82	0
2	A	807	0	844	25	0
2	E	807	0	844	13	0
3	B	646	0	687	17	0
3	F	694	0	742	33	0
4	C	812	0	866	32	0
4	G	803	0	853	27	0
5	D	751	0	779	16	0
5	H	730	0	753	12	0
6	A	23	0	0	2	0
6	B	16	0	0	4	0
6	C	29	0	0	8	0
6	D	28	0	0	1	0
6	E	41	0	0	2	0
6	F	26	0	0	12	0
6	G	26	0	0	1	0
6	H	11	0	0	0	0
6	I	32	0	0	16	0
6	J	32	0	0	11	0
All	All	12294	0	9670	299	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 14.

The worst 5 of 299 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:F:292:ARG:CB	6:F:517:HOH:O	1.65	1.32
1:J:166:DT:C5'	6:J:529:HOH:O	1.65	1.28

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:I:28:DA:O3'	6:I:522:HOH:O	1.63	1.12
3:F:292:ARG:HB2	6:F:517:HOH:O	1.23	1.12
3:B:35:ARG:HB2	6:B:487:HOH:O	1.55	1.03

There are no symmetry-related clashes.

## 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	96/135 (71%)	95 (99%)	1 (1%)	0	100	100
2	E	96/135 (71%)	95 (99%)	0	1 (1%)	19	21
3	B	79/103 (77%)	77 (98%)	2 (2%)	0	100	100
3	F	84/103 (82%)	79 (94%)	2 (2%)	3 (4%)	4	2
4	C	104/123 (85%)	100 (96%)	3 (3%)	1 (1%)	19	21
4	G	103/123 (84%)	101 (98%)	2 (2%)	0	100	100
5	D	93/123 (76%)	92 (99%)	0	1 (1%)	17	18
5	H	91/123 (74%)	88 (97%)	0	3 (3%)	5	3
All	All	746/968 (77%)	727 (98%)	10 (1%)	9 (1%)	16	16

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
5	D	1301	GLY
2	E	734	ARG
3	F	219	ARG
3	F	220	LYS
5	H	1430	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	85/110 (77%)	81 (95%)	4 (5%)	32	43
2	E	85/110 (77%)	85 (100%)	0	100	100
3	B	66/79 (84%)	65 (98%)	1 (2%)	72	85
3	F	71/79 (90%)	70 (99%)	1 (1%)	74	86
4	C	82/93 (88%)	78 (95%)	4 (5%)	31	41
4	G	81/93 (87%)	79 (98%)	2 (2%)	55	73
5	D	82/104 (79%)	81 (99%)	1 (1%)	78	89
5	H	80/104 (77%)	77 (96%)	3 (4%)	40	54
All	All	632/772 (82%)	616 (98%)	16 (2%)	55	73

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	C	918	LYS
4	C	919	LYS
4	G	1088	ARG
4	C	881	ARG
5	H	1433	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
2	E	668	GLN
5	H	1492	GLN
4	G	1019	ASN
4	C	831	HIS
4	G	1031	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 carbohydrates 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	I	146/146 (100%)	1.43	42 (28%) 1 1	42, 98, 166, 185	0
1	J	146/146 (100%)	1.44	41 (28%) 1 1	49, 102, 158, 200	0
2	A	98/135 (72%)	0.84	3 (3%) 52 62	30, 48, 73, 124	0
2	E	98/135 (72%)	0.93	7 (7%) 19 26	27, 39, 65, 130	0
3	B	81/103 (78%)	0.92	8 (9%) 9 14	31, 44, 71, 161	0
3	F	86/103 (83%)	1.40	15 (17%) 2 3	25, 39, 70, 163	0
4	C	106/123 (86%)	0.89	9 (8%) 13 19	28, 42, 78, 150	0
4	G	105/123 (85%)	0.74	6 (5%) 27 36	35, 48, 89, 138	0
5	D	95/123 (77%)	1.01	12 (12%) 5 8	28, 41, 83, 119	0
5	H	93/123 (75%)	1.00	12 (12%) 5 7	32, 46, 86, 133	0
All	All	1054/1260 (83%)	1.09	155 (14%) 3 5	25, 51, 135, 200	0

The worst 5 of 155 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	J	273	DA	10.1
2	A	535	ALA	8.1
3	B	23	ARG	7.9
4	G	1014	ALA	7.8
1	J	272	DA	7.6

### 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 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.