



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 15, 2017 – 11:38 AM EDT

PDB ID : 2HII  
Title : heterotrimeric PCNA sliding clamp  
Authors : Pascal, J.M.; Tsodikov, O.V.; Ellenberger, T.  
Deposited on : unknown  
Resolution : 2.79 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
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030345
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	rb-20030345

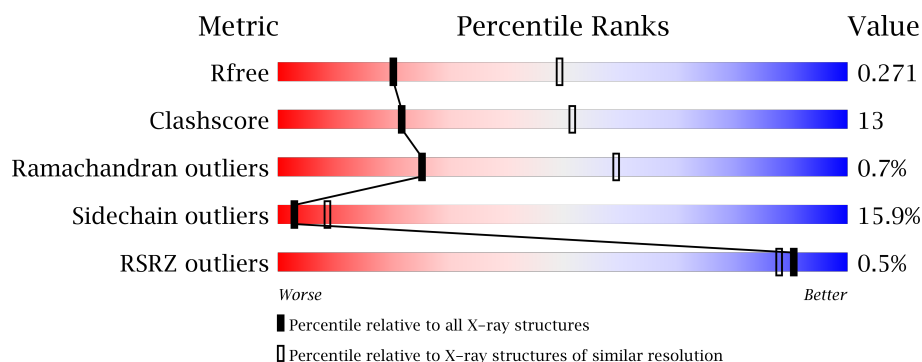
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.79 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))
$R_{free}$	100719	2583 (2.80-2.80)
Clashscore	112137	3033 (2.80-2.80)
Ramachandran outliers	110173	2983 (2.80-2.80)
Sidechain outliers	110143	2985 (2.80-2.80)
RSRZ outliers	101464	2610 (2.80-2.80)

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	A	257	
1	X	257	
2	B	245	
2	Y	245	
3	C	252	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	Z	252	 A horizontal bar chart showing the quality of chain Z. The bar is divided into four segments: green (64%), yellow (27%), orange (5%), and grey (5%). The percentages are labeled below the bar segments. A small black dot is visible at the end of the grey segment.

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 11550 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 PCNA1 (SSO0397).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	249	Total	C	N	O	Se	0	0	0
			1927	1228	309	381	9			
1	X	249	Total	C	N	O	Se	0	0	0
			1927	1228	309	381	9			

There are 34 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	MET	MODIFIED RESIDUE	UNP P57766
A	47	MSE	MET	MODIFIED RESIDUE	UNP P57766
A	50	MSE	MET	MODIFIED RESIDUE	UNP P57766
A	157	MSE	MET	MODIFIED RESIDUE	UNP P57766
A	182	MSE	MET	MODIFIED RESIDUE	UNP P57766
A	204	MSE	MET	MODIFIED RESIDUE	UNP P57766
A	220	MSE	MET	MODIFIED RESIDUE	UNP P57766
A	229	MSE	MET	MODIFIED RESIDUE	UNP P57766
A	241	MSE	MET	MODIFIED RESIDUE	UNP P57766
A	250	LEU	-	CLONING ARTIFACT	UNP P57766
A	251	GLU	-	CLONING ARTIFACT	UNP P57766
A	252	HIS	-	EXPRESSION TAG	UNP P57766
A	253	HIS	-	EXPRESSION TAG	UNP P57766
A	254	HIS	-	EXPRESSION TAG	UNP P57766
A	255	HIS	-	EXPRESSION TAG	UNP P57766
A	256	HIS	-	EXPRESSION TAG	UNP P57766
A	257	HIS	-	EXPRESSION TAG	UNP P57766
X	1	MSE	MET	MODIFIED RESIDUE	UNP P57766
X	47	MSE	MET	MODIFIED RESIDUE	UNP P57766
X	50	MSE	MET	MODIFIED RESIDUE	UNP P57766
X	157	MSE	MET	MODIFIED RESIDUE	UNP P57766
X	182	MSE	MET	MODIFIED RESIDUE	UNP P57766
X	204	MSE	MET	MODIFIED RESIDUE	UNP P57766
X	220	MSE	MET	MODIFIED RESIDUE	UNP P57766
X	229	MSE	MET	MODIFIED RESIDUE	UNP P57766

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
X	241	MSE	MET	MODIFIED RESIDUE	UNP P57766
X	250	LEU	-	CLONING ARTIFACT	UNP P57766
X	251	GLU	-	CLONING ARTIFACT	UNP P57766
X	252	HIS	-	EXPRESSION TAG	UNP P57766
X	253	HIS	-	EXPRESSION TAG	UNP P57766
X	254	HIS	-	EXPRESSION TAG	UNP P57766
X	255	HIS	-	EXPRESSION TAG	UNP P57766
X	256	HIS	-	EXPRESSION TAG	UNP P57766
X	257	HIS	-	EXPRESSION TAG	UNP P57766

- Molecule 2 is a protein called PCNA2 (SSO1047).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	243	Total	C	N	O	Se	0	0	0
			1922	1237	301	380	4			
2	Y	243	Total	C	N	O	Se	0	0	0
			1922	1237	301	380	4			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	2	MSE	MET	MODIFIED RESIDUE	UNP Q97Z84
B	199	MSE	MET	MODIFIED RESIDUE	UNP Q97Z84
B	208	MSE	MET	MODIFIED RESIDUE	UNP Q97Z84
B	215	MSE	MET	MODIFIED RESIDUE	UNP Q97Z84
Y	2	MSE	MET	MODIFIED RESIDUE	UNP Q97Z84
Y	199	MSE	MET	MODIFIED RESIDUE	UNP Q97Z84
Y	208	MSE	MET	MODIFIED RESIDUE	UNP Q97Z84
Y	215	MSE	MET	MODIFIED RESIDUE	UNP Q97Z84

- Molecule 3 is a protein called PCNA3 (SSO0405).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	243	Total	C	N	O	Se	0	0	0
			1926	1224	311	387	4			
3	Z	243	Total	C	N	O	Se	0	0	0
			1926	1224	311	387	4			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	1	MSE	MET	MODIFIED RESIDUE	UNP P57765

*Continued on next page...*

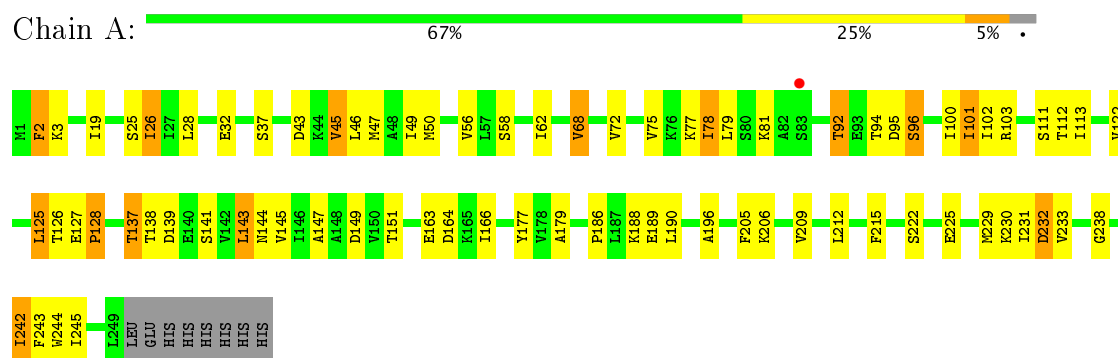
*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	55	MSE	MET	MODIFIED RESIDUE	UNP P57765
C	75	MSE	MET	MODIFIED RESIDUE	UNP P57765
C	230	MSE	MET	MODIFIED RESIDUE	UNP P57765
C	245	LEU	-	CLONING ARTIFACT	UNP P57765
C	246	GLU	-	CLONING ARTIFACT	UNP P57765
C	247	HIS	-	EXPRESSION TAG	UNP P57765
C	248	HIS	-	EXPRESSION TAG	UNP P57765
C	249	HIS	-	EXPRESSION TAG	UNP P57765
C	250	HIS	-	EXPRESSION TAG	UNP P57765
C	251	HIS	-	EXPRESSION TAG	UNP P57765
C	252	HIS	-	EXPRESSION TAG	UNP P57765
Z	1	MSE	MET	MODIFIED RESIDUE	UNP P57765
Z	55	MSE	MET	MODIFIED RESIDUE	UNP P57765
Z	75	MSE	MET	MODIFIED RESIDUE	UNP P57765
Z	230	MSE	MET	MODIFIED RESIDUE	UNP P57765
Z	245	LEU	-	CLONING ARTIFACT	UNP P57765
Z	246	GLU	-	CLONING ARTIFACT	UNP P57765
Z	247	HIS	-	EXPRESSION TAG	UNP P57765
Z	248	HIS	-	EXPRESSION TAG	UNP P57765
Z	249	HIS	-	EXPRESSION TAG	UNP P57765
Z	250	HIS	-	EXPRESSION TAG	UNP P57765
Z	251	HIS	-	EXPRESSION TAG	UNP P57765
Z	252	HIS	-	EXPRESSION TAG	UNP P57765

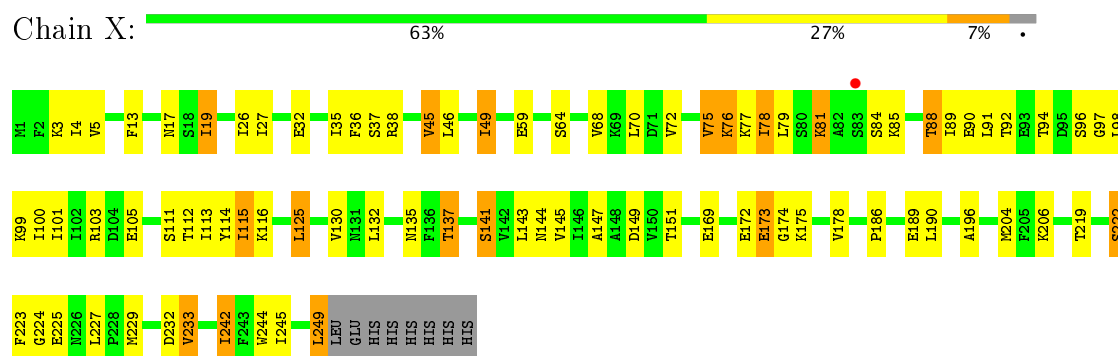
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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.

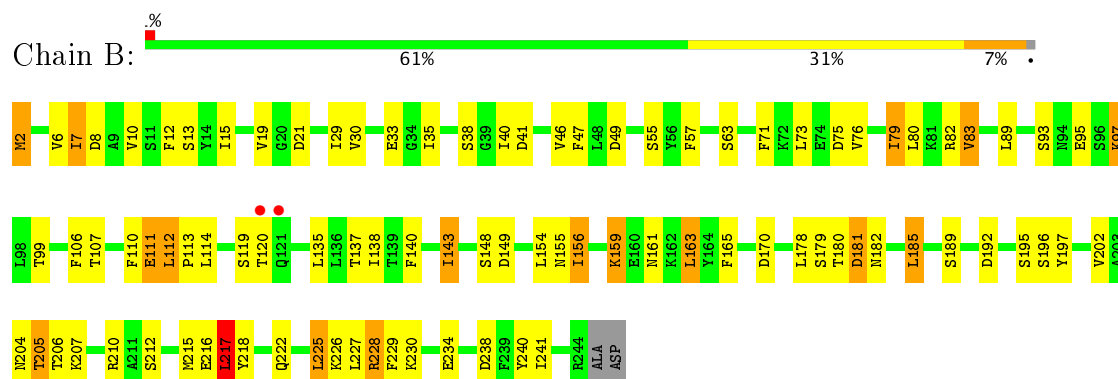
- Molecule 1: PCNA1 (SSO0397)



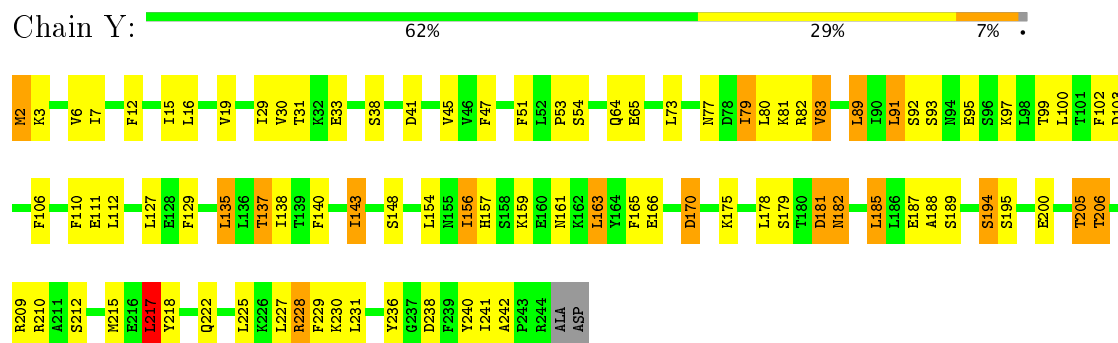
- Molecule 1: PCNA1 (SSO0397)



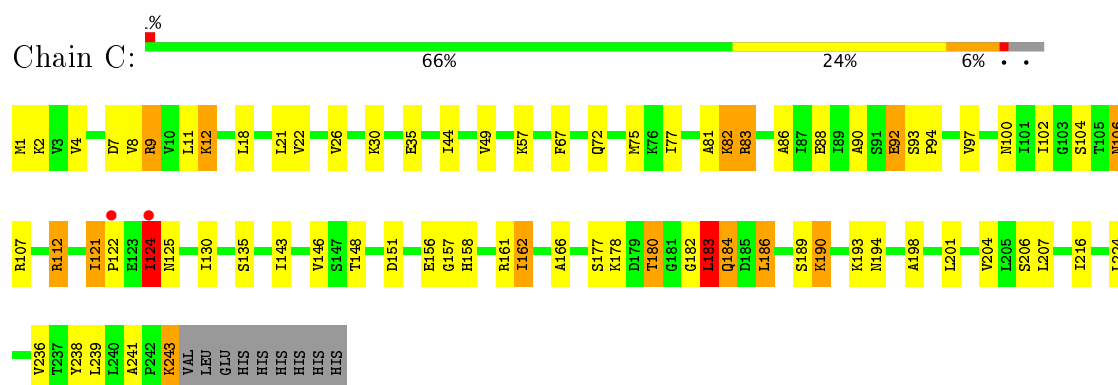
- Molecule 2: PCNA2 (SSO1047)



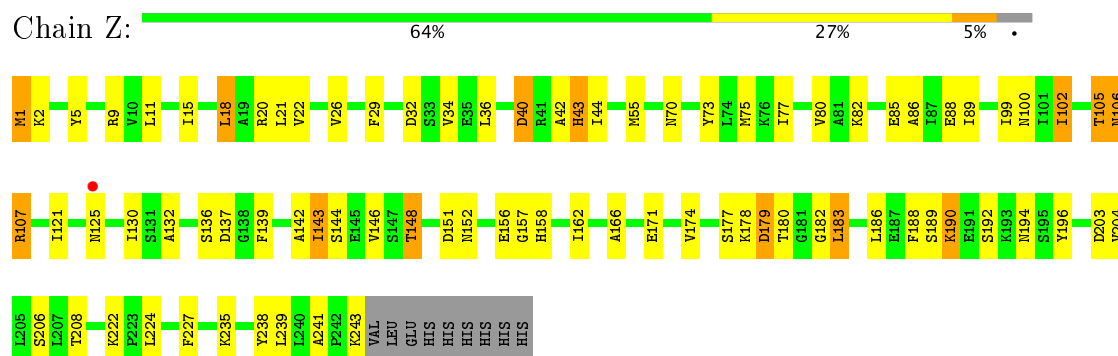
- Molecule 2: PCNA2 (SSO1047)



- Molecule 3: PCNA3 (SSO0405)



- Molecule 3: PCNA3 (SSO0405)





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	36.52Å 87.18Å 135.87Å 89.80° 86.36° 78.33°	Depositor
Resolution (Å)	20.00 – 2.79 45.20 – 2.79	Depositor EDS
% Data completeness (in resolution range)	92.8 (20.00-2.79) 92.7 (45.20-2.79)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.56 (at 2.81Å)	Xtriage
Refinement program	REFMAC	Depositor
R, $R_{free}$	0.215 , 0.282 0.206 , 0.271	Depositor DCC
$R_{free}$ test set	1892 reflections (5.30%)	DCC
Wilson B-factor (Å <sup>2</sup> )	39.9	Xtriage
Anisotropy	0.906	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 11.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.000 for -h,-h+k,-l	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	11550	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 26.04 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 2.8350e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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	0.53	0/1947	0.68	1/2610 (0.0%)
1	X	0.51	0/1947	0.68	0/2610
2	B	0.47	0/1954	0.68	2/2636 (0.1%)
2	Y	0.55	0/1954	0.73	2/2636 (0.1%)
3	C	0.53	0/1949	0.68	1/2622 (0.0%)
3	Z	0.51	0/1949	0.65	0/2622
All	All	0.52	0/11700	0.68	6/15736 (0.0%)

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	217	LEU	CA-CB-CG	5.94	128.95	115.30
1	A	125	LEU	CA-CB-CG	5.73	128.47	115.30
2	Y	217	LEU	CA-CB-CG	5.37	127.64	115.30
2	Y	91	LEU	CA-CB-CG	5.31	127.50	115.30
2	B	135	LEU	CA-CB-CG	5.13	127.11	115.30

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	A	1927	0	1976	50	0
1	X	1927	0	1976	57	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	1922	0	1920	58	0
2	Y	1922	0	1920	60	0
3	C	1926	0	1936	53	0
3	Z	1926	0	1936	57	0
All	All	11550	0	11664	311	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 13.

The worst 5 of 311 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)
1:X:149:ASP:HB2	2:Y:110:PHE:HZ	1.13	1.13
2:Y:215:MSE:HE2	2:Y:229:PHE:CE2	1.89	1.06
1:X:149:ASP:HB3	2:Y:82:ARG:HE	1.16	1.05
2:Y:137:THR:HB	2:Y:215:MSE:HE3	1.33	1.03
2:Y:80:LEU:O	2:Y:83:VAL:HG22	1.61	0.99

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	
1	A	247/257 (96%)	237 (96%)	9 (4%)	1 (0%)	38	72
1	X	247/257 (96%)	238 (96%)	8 (3%)	1 (0%)	38	72
2	B	241/245 (98%)	228 (95%)	11 (5%)	2 (1%)	22	55
2	Y	241/245 (98%)	231 (96%)	8 (3%)	2 (1%)	22	55
3	C	241/252 (96%)	235 (98%)	4 (2%)	2 (1%)	22	55
3	Z	241/252 (96%)	233 (97%)	6 (2%)	2 (1%)	22	55

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	1458/1508 (97%)	1402 (96%)	46 (3%)	10 (1%)	25 59

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	C	124	ILE
2	B	148	SER
3	C	190	LYS
2	Y	182	ASN
3	Z	190	LYS

### 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
1	A	219/219 (100%)	190 (87%)	29 (13%)	5 14
1	X	219/219 (100%)	182 (83%)	37 (17%)	2 7
2	B	218/215 (101%)	178 (82%)	40 (18%)	2 6
2	Y	218/215 (101%)	177 (81%)	41 (19%)	2 5
3	C	219/224 (98%)	187 (85%)	32 (15%)	3 11
3	Z	219/224 (98%)	189 (86%)	30 (14%)	4 13
All	All	1312/1316 (100%)	1103 (84%)	209 (16%)	3 9

5 of 209 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	C	183	LEU
1	X	85	LYS
3	Z	105	THR
3	C	194	ASN
1	X	49	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 21 such sidechains are listed below:

Mol	Chain	Res	Type
2	Y	27	ASN
2	Y	121	GLN
3	Z	106	ASN
1	X	240	HIS
3	Z	119	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 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	A	240/257 (93%)	-0.08	1 (0%) 92 90	28, 32, 37, 44	0
1	X	240/257 (93%)	0.00	1 (0%) 92 90	27, 32, 36, 40	0
2	B	239/245 (97%)	-0.04	2 (0%) 86 81	28, 32, 36, 42	0
2	Y	239/245 (97%)	-0.05	0 100 100	28, 32, 35, 40	0
3	C	239/252 (94%)	-0.05	2 (0%) 86 81	28, 32, 37, 49	0
3	Z	239/252 (94%)	-0.05	1 (0%) 92 90	28, 32, 36, 46	0
All	All	1436/1508 (95%)	-0.05	7 (0%) 90 88	27, 32, 36, 49	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	X	83	SER	4.7
2	B	121	GLN	2.9
2	B	120	THR	2.8
1	A	83	SER	2.5
3	C	122	PRO	2.3

### 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.