



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 06:21 am GMT

PDB ID : 2GWF  
Title : Structure of a USP8-NRDP1 complex  
Authors : Walker, J.R.; Avvakumov, G.V.; Xue, S.; Newman, E.M.; Butler-Cole, C.; Finerty Jr., P.J.; Weigelt, J.; Sundstrom, M.; Arrowsmith, C.H.; Edwards, A.M.; Bochkarev, A.; Dhe-Paganon, S.; Structural Genomics Consortium (SGC)  
Deposited on : 2006-05-04  
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
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
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)	:	recalc28949

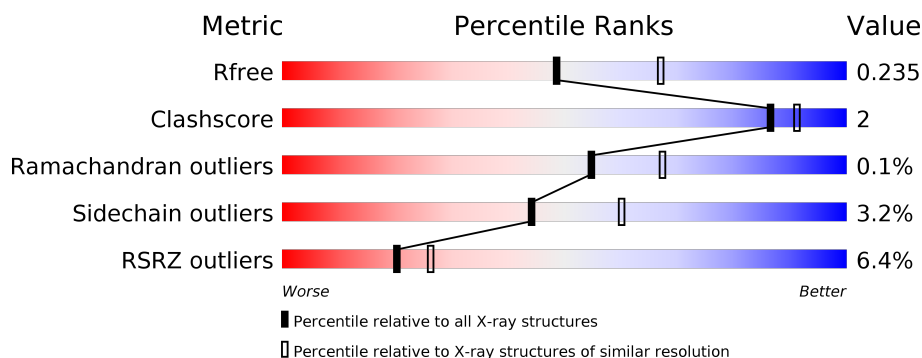
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.30 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))
$R_{free}$	100719	4130 (2.30-2.30)
Clashscore	112137	4751 (2.30-2.30)
Ramachandran outliers	110173	4705 (2.30-2.30)
Sidechain outliers	110143	4704 (2.30-2.30)
RSRZ outliers	101464	4156 (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	A	157	<div> <div>4%</div> <div> <div></div> <div>79%</div> <div>6%</div> <div>13%</div> </div> </div>
1	C	157	<div> <div>8%</div> <div> <div></div> <div>76%</div> <div>8%</div> <div>15%</div> </div> </div>
1	E	157	<div> <div>6%</div> <div> <div></div> <div>81%</div> <div>5%</div> <div>13%</div> </div> </div>
2	B	134	<div> <div>3%</div> <div> <div></div> <div>83%</div> <div>7%</div> <div>10%</div> </div> </div>
2	D	134	<div> <div>7%</div> <div> <div></div> <div>91%</div> <div>•</div> <div>6%</div> </div> </div>
2	F	134	<div> <div>7%</div> <div> <div></div> <div>81%</div> <div>9%</div> <div>9%</div> </div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 6388 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 Ubiquitin carboxyl-terminal hydrolase 8.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	136	Total	C	N	O	S	0	0	0
			1088	693	178	211	6			
1	C	134	Total	C	N	O	S	0	0	0
			1075	686	175	208	6			
1	E	137	Total	C	N	O	S	0	0	0
			1097	699	180	212	6			

There are 57 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	162	MET	-	INITIATING METHIONINE	UNP P40818
A	163	GLY	-	CLONING ARTIFACT	UNP P40818
A	164	SER	-	CLONING ARTIFACT	UNP P40818
A	165	SER	-	CLONING ARTIFACT	UNP P40818
A	166	HIS	-	EXPRESSION TAG	UNP P40818
A	167	HIS	-	EXPRESSION TAG	UNP P40818
A	168	HIS	-	EXPRESSION TAG	UNP P40818
A	169	HIS	-	EXPRESSION TAG	UNP P40818
A	170	HIS	-	EXPRESSION TAG	UNP P40818
A	171	HIS	-	EXPRESSION TAG	UNP P40818
A	172	SER	-	CLONING ARTIFACT	UNP P40818
A	173	SER	-	CLONING ARTIFACT	UNP P40818
A	174	GLY	-	CLONING ARTIFACT	UNP P40818
A	175	LEU	-	CLONING ARTIFACT	UNP P40818
A	176	VAL	-	CLONING ARTIFACT	UNP P40818
A	177	PRO	-	CLONING ARTIFACT	UNP P40818
A	178	ARG	-	CLONING ARTIFACT	UNP P40818
A	179	GLY	-	CLONING ARTIFACT	UNP P40818
A	180	SER	-	CLONING ARTIFACT	UNP P40818
C	162	MET	-	INITIATING METHIONINE	UNP P40818
C	163	GLY	-	CLONING ARTIFACT	UNP P40818
C	164	SER	-	CLONING ARTIFACT	UNP P40818
C	165	SER	-	CLONING ARTIFACT	UNP P40818

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	166	HIS	-	EXPRESSION TAG	UNP P40818
C	167	HIS	-	EXPRESSION TAG	UNP P40818
C	168	HIS	-	EXPRESSION TAG	UNP P40818
C	169	HIS	-	EXPRESSION TAG	UNP P40818
C	170	HIS	-	EXPRESSION TAG	UNP P40818
C	171	HIS	-	EXPRESSION TAG	UNP P40818
C	172	SER	-	CLONING ARTIFACT	UNP P40818
C	173	SER	-	CLONING ARTIFACT	UNP P40818
C	174	GLY	-	CLONING ARTIFACT	UNP P40818
C	175	LEU	-	CLONING ARTIFACT	UNP P40818
C	176	VAL	-	CLONING ARTIFACT	UNP P40818
C	177	PRO	-	CLONING ARTIFACT	UNP P40818
C	178	ARG	-	CLONING ARTIFACT	UNP P40818
C	179	GLY	-	CLONING ARTIFACT	UNP P40818
C	180	SER	-	CLONING ARTIFACT	UNP P40818
E	162	MET	-	INITIATING METHIONINE	UNP P40818
E	163	GLY	-	CLONING ARTIFACT	UNP P40818
E	164	SER	-	CLONING ARTIFACT	UNP P40818
E	165	SER	-	CLONING ARTIFACT	UNP P40818
E	166	HIS	-	EXPRESSION TAG	UNP P40818
E	167	HIS	-	EXPRESSION TAG	UNP P40818
E	168	HIS	-	EXPRESSION TAG	UNP P40818
E	169	HIS	-	EXPRESSION TAG	UNP P40818
E	170	HIS	-	EXPRESSION TAG	UNP P40818
E	171	HIS	-	EXPRESSION TAG	UNP P40818
E	172	SER	-	CLONING ARTIFACT	UNP P40818
E	173	SER	-	CLONING ARTIFACT	UNP P40818
E	174	GLY	-	CLONING ARTIFACT	UNP P40818
E	175	LEU	-	CLONING ARTIFACT	UNP P40818
E	176	VAL	-	CLONING ARTIFACT	UNP P40818
E	177	PRO	-	CLONING ARTIFACT	UNP P40818
E	178	ARG	-	CLONING ARTIFACT	UNP P40818
E	179	GLY	-	CLONING ARTIFACT	UNP P40818
E	180	SER	-	CLONING ARTIFACT	UNP P40818

- Molecule 2 is a protein called RING finger protein 41.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	121	Total	C	N	O	S	0	0	0
			952	595	171	178	8			
2	D	126	Total	C	N	O	S	0	0	0
			994	622	176	188	8			

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	F	122	Total 964	C 604	N 172	O 180	S 8	0	0	0

There are 27 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	184	SER	-	CLONING ARTIFACT	UNP Q9H4P4
B	185	SER	-	CLONING ARTIFACT	UNP Q9H4P4
B	186	GLY	-	CLONING ARTIFACT	UNP Q9H4P4
B	187	LEU	-	CLONING ARTIFACT	UNP Q9H4P4
B	188	VAL	-	CLONING ARTIFACT	UNP Q9H4P4
B	189	PRO	-	CLONING ARTIFACT	UNP Q9H4P4
B	190	ARG	-	CLONING ARTIFACT	UNP Q9H4P4
B	191	GLY	-	CLONING ARTIFACT	UNP Q9H4P4
B	192	SER	-	CLONING ARTIFACT	UNP Q9H4P4
D	184	SER	-	CLONING ARTIFACT	UNP Q9H4P4
D	185	SER	-	CLONING ARTIFACT	UNP Q9H4P4
D	186	GLY	-	CLONING ARTIFACT	UNP Q9H4P4
D	187	LEU	-	CLONING ARTIFACT	UNP Q9H4P4
D	188	VAL	-	CLONING ARTIFACT	UNP Q9H4P4
D	189	PRO	-	CLONING ARTIFACT	UNP Q9H4P4
D	190	ARG	-	CLONING ARTIFACT	UNP Q9H4P4
D	191	GLY	-	CLONING ARTIFACT	UNP Q9H4P4
D	192	SER	-	CLONING ARTIFACT	UNP Q9H4P4
F	184	SER	-	CLONING ARTIFACT	UNP Q9H4P4
F	185	SER	-	CLONING ARTIFACT	UNP Q9H4P4
F	186	GLY	-	CLONING ARTIFACT	UNP Q9H4P4
F	187	LEU	-	CLONING ARTIFACT	UNP Q9H4P4
F	188	VAL	-	CLONING ARTIFACT	UNP Q9H4P4
F	189	PRO	-	CLONING ARTIFACT	UNP Q9H4P4
F	190	ARG	-	CLONING ARTIFACT	UNP Q9H4P4
F	191	GLY	-	CLONING ARTIFACT	UNP Q9H4P4
F	192	SER	-	CLONING ARTIFACT	UNP Q9H4P4

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	45	Total 45	O 45	0	0
3	B	29	Total 29	O 29	0	0

*Continued on next page...*

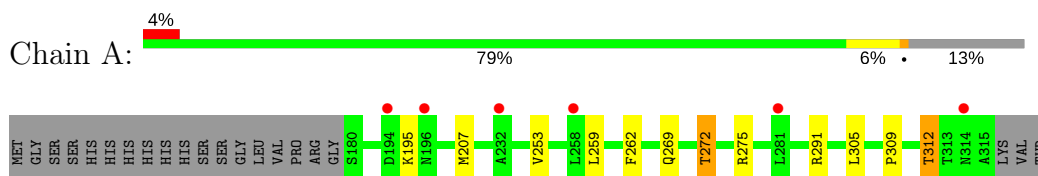
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	C	45	Total 45	O 45	0	0
3	D	30	Total 30	O 30	0	0
3	E	37	Total 37	O 37	0	0
3	F	32	Total 32	O 32	0	0

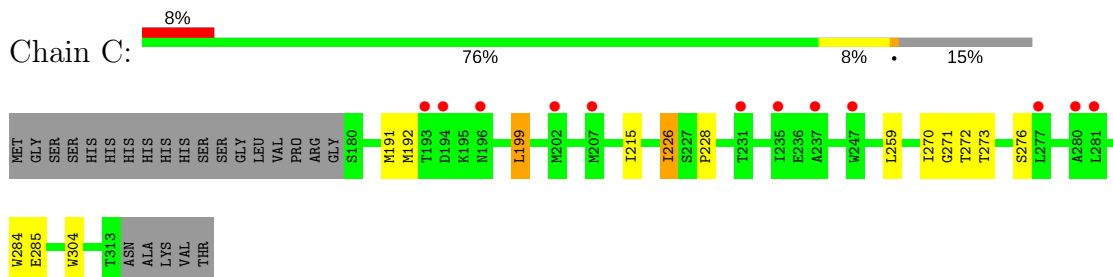
### 3 Residue-property plots [i](#)

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.

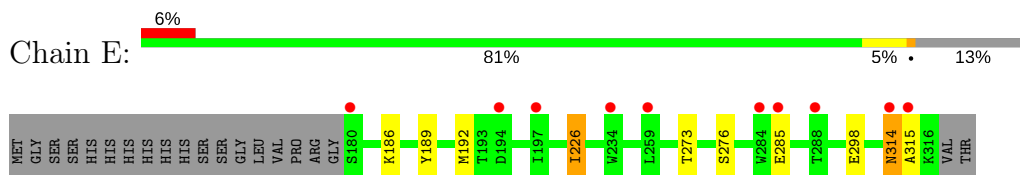
- Molecule 1: Ubiquitin carboxyl-terminal hydrolase 8



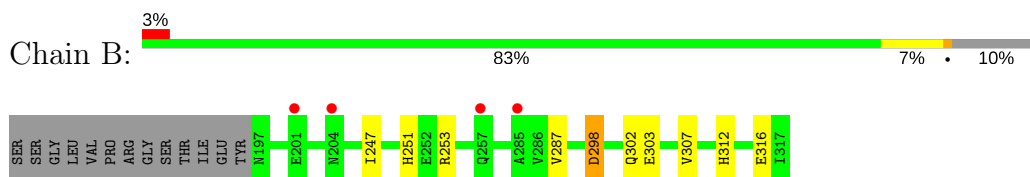
- Molecule 1: Ubiquitin carboxyl-terminal hydrolase 8



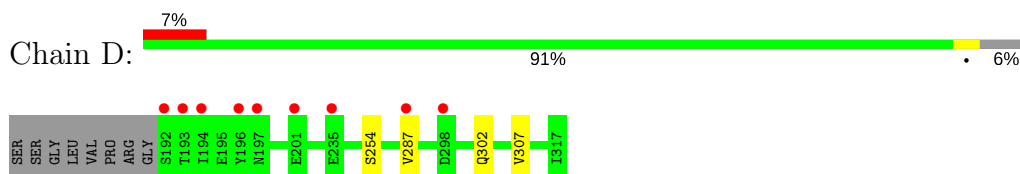
- Molecule 1: Ubiquitin carboxyl-terminal hydrolase 8



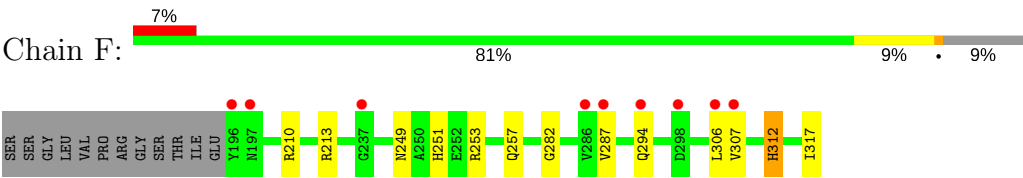
- Molecule 2: RING finger protein 41



- Molecule 2: RING finger protein 41



● Molecule 2: RING finger protein 41





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	76.11Å 67.09Å 99.28Å 90.00° 91.12° 90.00°	Depositor
Resolution (Å)	30.00 – 2.30 29.91 – 2.30	Depositor EDS
% Data completeness (in resolution range)	93.2 (30.00-2.30) 93.2 (29.91-2.30)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.45 (at 2.31Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.206 , 0.237 0.201 , 0.235	Depositor DCC
$R_{free}$ test set	2093 reflections (5.29%)	DCC
Wilson B-factor (Å <sup>2</sup> )	38.9	Xtriage
Anisotropy	0.131	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 47.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.019 for h,-k,-l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	6388	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.51% 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	0/1112	0.59	0/1511
1	C	0.34	0/1099	0.58	0/1493
1	E	0.33	0/1121	0.60	0/1522
2	B	0.34	0/970	0.60	0/1316
2	D	0.33	0/1013	0.56	0/1375
2	F	0.34	0/983	0.60	0/1334
All	All	0.33	0/6298	0.59	0/8551

There are no bond length outliers.

There are no bond angle outliers.

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	1088	0	1074	5	0
1	C	1075	0	1063	10	0
1	E	1097	0	1087	5	0
2	B	952	0	942	4	0
2	D	994	0	980	2	0
2	F	964	0	951	6	0
3	A	45	0	0	0	0
3	B	29	0	0	0	0
3	C	45	0	0	1	0
3	D	30	0	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	E	37	0	0	0	0
3	F	32	0	0	1	0
All	All	6388	0	6097	30	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 2.

The worst 5 of 30 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:F:249:ASN:HB3	2:F:306:LEU:HD12	1.67	0.76
1:E:226:ILE:HD11	1:E:273:THR:HA	1.68	0.75
2:F:251:HIS:HD2	2:F:253:ARG:H	1.35	0.73
1:C:226:ILE:HD12	1:C:276:SER:HB2	1.74	0.70
2:B:251:HIS:HD2	2:B:253:ARG:H	1.40	0.69

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	134/157 (85%)	130 (97%)	3 (2%)	1 (1%)	25	30
1	C	132/157 (84%)	128 (97%)	4 (3%)	0	100	100
1	E	135/157 (86%)	128 (95%)	7 (5%)	0	100	100
2	B	119/134 (89%)	117 (98%)	2 (2%)	0	100	100
2	D	124/134 (92%)	122 (98%)	2 (2%)	0	100	100
2	F	120/134 (90%)	118 (98%)	2 (2%)	0	100	100
All	All	764/873 (88%)	743 (97%)	20 (3%)	1 (0%)	55	67

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	195	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	121/139 (87%)	116 (96%)	5 (4%)	35	48
1	C	120/139 (86%)	116 (97%)	4 (3%)	43	59
1	E	122/139 (88%)	117 (96%)	5 (4%)	35	48
2	B	103/114 (90%)	98 (95%)	5 (5%)	29	39
2	D	108/114 (95%)	108 (100%)	0	100	100
2	F	104/114 (91%)	101 (97%)	3 (3%)	48	64
All	All	678/759 (89%)	656 (97%)	22 (3%)	44	60

5 of 22 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	312	HIS
1	C	259	LEU
2	F	294	GLN
1	C	199	LEU
1	C	226	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	257	GLN
2	D	244	ASN
2	F	251	HIS
2	B	251	HIS
2	F	244	ASN

### 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	136/157 (86%)	0.42	6 (4%) 35 42	18, 26, 39, 44	0
1	C	134/157 (85%)	0.50	12 (8%) 10 14	20, 26, 41, 53	0
1	E	137/157 (87%)	0.63	10 (7%) 16 21	20, 33, 49, 53	0
2	B	121/134 (90%)	0.30	4 (3%) 47 54	17, 29, 39, 43	0
2	D	126/134 (94%)	0.62	9 (7%) 17 22	20, 28, 50, 74	0
2	F	122/134 (91%)	0.40	9 (7%) 15 21	16, 26, 35, 43	0
All	All	776/873 (88%)	0.48	50 (6%) 20 26	16, 28, 44, 74	0

The worst 5 of 50 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	194	ILE	6.5
1	E	285	GLU	6.2
2	D	192	SER	6.0
1	C	194	ASP	5.3
2	F	196	TYR	5.2

### 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.