



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 01:28 am GMT

PDB ID : 5I44  
Title : Structure of RacA-DNA complex; P21 form  
Authors : Schumacher, M.A.  
Deposited on : 2016-02-11  
Resolution : 2.62 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

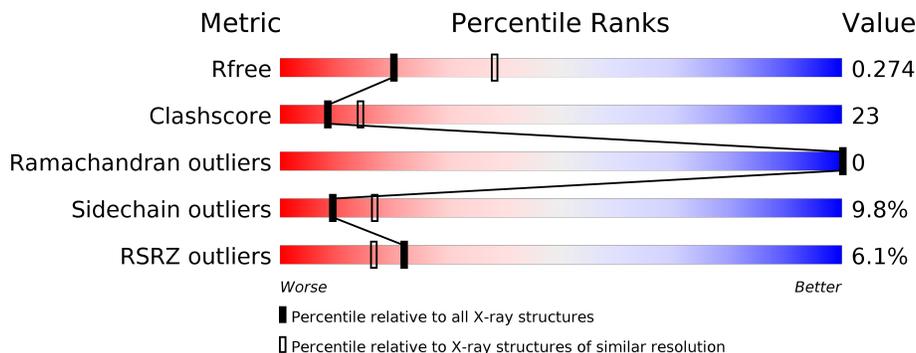
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.62 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	2983 (2.64-2.60)
Clashscore	112137	3351 (2.64-2.60)
Ramachandran outliers	110173	3298 (2.64-2.60)
Sidechain outliers	110143	3298 (2.64-2.60)
RSRZ outliers	101464	2992 (2.64-2.60)

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	69	
1	B	69	
1	D	69	
1	E	69	
1	F	69	
1	G	69	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	H	69	
1	I	69	
1	J	69	
1	K	69	
2	P	14	
2	R	14	
2	T	14	
2	U	14	
2	W	14	
2	Z	14	

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 7079 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 Chromosome-anchoring protein RacA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	B	68	528	330	95	101	2	0	0	0
1	A	67	521	325	94	100	2	0	0	0
1	D	67	521	325	94	100	2	0	0	0
1	E	68	528	330	95	101	2	0	0	0
1	G	68	534	334	96	102	2	0	0	0
1	F	68	528	330	95	101	2	0	0	0
1	H	64	501	314	89	96	2	0	0	0
1	I	65	508	319	90	97	2	0	0	0
1	J	67	521	325	94	100	2	0	0	0
1	K	66	517	323	93	99	2	0	0	0

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	2	GLY	-	expression tag	UNP P45870
B	3	SER	-	expression tag	UNP P45870
B	4	HIS	-	expression tag	UNP P45870
B	54	LYS	GLN	conflict	UNP P45870
A	2	GLY	-	expression tag	UNP P45870
A	3	SER	-	expression tag	UNP P45870
A	4	HIS	-	expression tag	UNP P45870
A	54	LYS	GLN	conflict	UNP P45870
D	2	GLY	-	expression tag	UNP P45870

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
D	3	SER	-	expression tag	UNP P45870
D	4	HIS	-	expression tag	UNP P45870
D	54	LYS	GLN	conflict	UNP P45870
E	2	GLY	-	expression tag	UNP P45870
E	3	SER	-	expression tag	UNP P45870
E	4	HIS	-	expression tag	UNP P45870
E	54	LYS	GLN	conflict	UNP P45870
G	0	GLY	-	expression tag	UNP P45870
G	1	SER	-	expression tag	UNP P45870
G	2	HIS	-	expression tag	UNP P45870
G	52	LYS	GLN	conflict	UNP P45870
F	2	GLY	-	expression tag	UNP P45870
F	3	SER	-	expression tag	UNP P45870
F	4	HIS	-	expression tag	UNP P45870
F	54	LYS	GLN	conflict	UNP P45870
H	-2	GLY	-	expression tag	UNP P45870
H	-1	SER	-	expression tag	UNP P45870
H	0	HIS	-	expression tag	UNP P45870
H	50	LYS	GLN	conflict	UNP P45870
I	-2	GLY	-	expression tag	UNP P45870
I	-1	SER	-	expression tag	UNP P45870
I	0	HIS	-	expression tag	UNP P45870
I	50	LYS	GLN	conflict	UNP P45870
J	2	GLY	-	expression tag	UNP P45870
J	3	SER	-	expression tag	UNP P45870
J	4	HIS	-	expression tag	UNP P45870
J	54	LYS	GLN	conflict	UNP P45870
K	0	GLY	-	expression tag	UNP P45870
K	1	SER	-	expression tag	UNP P45870
K	2	HIS	-	expression tag	UNP P45870
K	52	LYS	GLN	conflict	UNP P45870

- Molecule 2 is a DNA chain called DNA (5'-D(\*TP\*GP\*AP\*CP\*GP\*CP\*CP\*GP\*GP\*CP\*GP\*TP\*CP\*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	U	14	Total	C	N	O	P	0	0	0
			284	135	54	82	13			
2	T	14	Total	C	N	O	P	0	0	0
			284	135	54	82	13			
2	Z	14	Total	C	N	O	P	0	0	0
			284	135	54	82	13			

*Continued on next page...*

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	R	14	Total 284	C 135	N 54	O 82	P 13	0	0	0
2	P	14	Total 284	C 135	N 54	O 82	P 13	0	0	0
2	W	14	Total 284	C 135	N 54	O 82	P 13	0	0	0

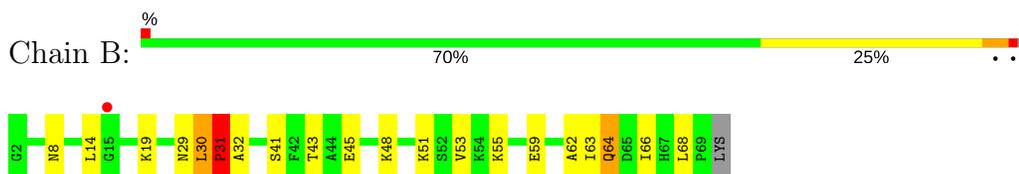
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	10	Total 10	O 10	0	0
3	A	16	Total 16	O 16	0	0
3	D	10	Total 10	O 10	0	0
3	E	14	Total 14	O 14	0	0
3	G	16	Total 16	O 16	0	0
3	F	12	Total 12	O 12	0	0
3	H	5	Total 5	O 5	0	0
3	I	7	Total 7	O 7	0	0
3	J	10	Total 10	O 10	0	0
3	K	14	Total 14	O 14	0	0
3	U	14	Total 14	O 14	0	0
3	T	12	Total 12	O 12	0	0
3	Z	4	Total 4	O 4	0	0
3	R	5	Total 5	O 5	0	0
3	P	9	Total 9	O 9	0	0
3	W	10	Total 10	O 10	0	0

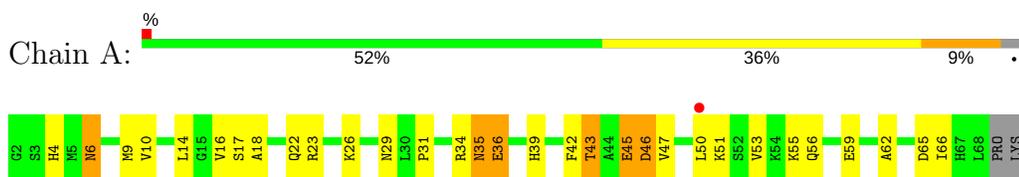
### 3 Residue-property plots [i](#)

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.

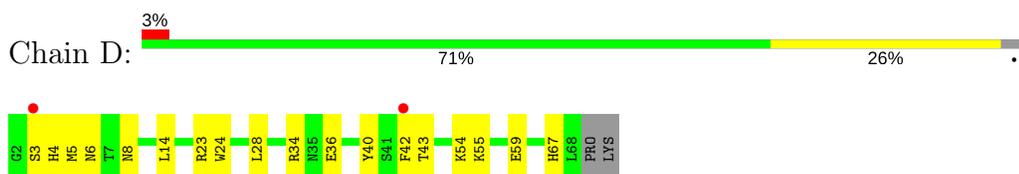
- Molecule 1: Chromosome-anchoring protein RacA



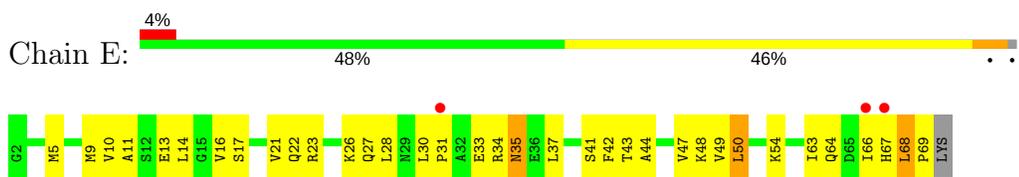
- Molecule 1: Chromosome-anchoring protein RacA



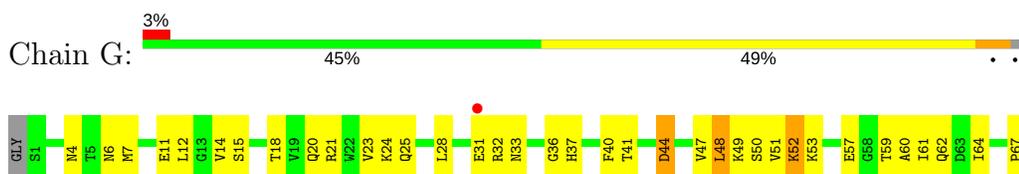
- Molecule 1: Chromosome-anchoring protein RacA



- Molecule 1: Chromosome-anchoring protein RacA



- Molecule 1: Chromosome-anchoring protein RacA

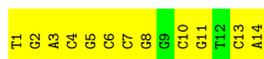


- Molecule 1: Chromosome-anchoring protein RacA



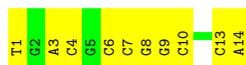


Chain Z:  14% 86%



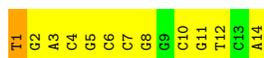
- Molecule 2: DNA (5'-D(\*TP\*GP\*AP\*CP\*GP\*CP\*CP\*GP\*GP\*CP\*GP\*TP\*CP\*A)-3')

Chain R:  29% 71%



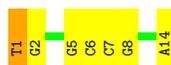
- Molecule 2: DNA (5'-D(\*TP\*GP\*AP\*CP\*GP\*CP\*CP\*GP\*GP\*CP\*GP\*TP\*CP\*A)-3')

Chain P:  14% 79% 7%



- Molecule 2: DNA (5'-D(\*TP\*GP\*AP\*CP\*GP\*CP\*CP\*GP\*GP\*CP\*GP\*TP\*CP\*A)-3')

Chain W:  50% 43% 7%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	56.60Å 68.50Å 117.40Å 90.00° 97.50° 90.00°	Depositor
Resolution (Å)	500.00 – 2.62 59.04 – 2.63	Depositor EDS
% Data completeness (in resolution range)	94.6 (500.00-2.62) 96.8 (59.04-2.63)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.93 (at 2.61Å)	Xtrriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.239 , 0.269 0.244 , 0.274	Depositor DCC
$R_{free}$ test set	1704 reflections (6.59%)	DCC
Wilson B-factor (Å <sup>2</sup> )	54.1	Xtrriage
Anisotropy	0.566	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 71.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	7079	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	68.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 18.18% 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	0.42	0/528	0.61	0/711
1	B	1.23	3/536 (0.6%)	1.07	5/723 (0.7%)
1	D	0.47	0/528	0.59	0/711
1	E	0.44	0/536	0.74	1/723 (0.1%)
1	F	0.49	0/536	0.68	1/723 (0.1%)
1	G	0.44	0/542	0.68	0/729
1	H	0.50	0/507	0.78	0/683
1	I	0.43	0/515	0.63	0/695
1	J	0.40	0/528	0.51	0/711
1	K	0.51	0/524	0.64	0/706
2	P	0.60	0/318	0.78	0/489
2	R	0.54	0/318	0.77	0/489
2	T	0.54	0/318	0.74	0/489
2	U	0.62	0/318	0.80	0/489
2	W	0.58	0/318	0.77	0/489
2	Z	0.48	0/318	0.69	0/489
All	All	0.58	3/7188 (0.0%)	0.72	7/10049 (0.1%)

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
2	P	0	1
2	T	0	1
2	W	0	1
All	All	0	3

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	32	ALA	N-CA	-18.98	1.08	1.46

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	31	PRO	N-CD	17.05	1.71	1.47
1	B	31	PRO	CA-CB	5.11	1.63	1.53

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	31	PRO	N-CA-CB	-11.66	89.31	103.30
1	B	31	PRO	C-N-CA	11.09	149.42	121.70
1	B	30	LEU	CB-CA-C	10.45	130.04	110.20
1	E	68	LEU	CA-CB-CG	7.97	133.62	115.30
1	B	32	ALA	N-CA-CB	7.62	120.76	110.10

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	P	1	DT	Sidechain
2	T	1	DT	Sidechain
2	W	1	DT	Sidechain

## 5.2 Too-close contacts

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	521	0	528	23	0
1	B	528	0	534	16	0
1	D	521	0	528	8	0
1	E	528	0	535	23	0
1	F	528	0	535	21	0
1	G	534	0	548	28	0
1	H	501	0	516	40	0
1	I	508	0	523	31	0
1	J	521	0	528	25	0
1	K	517	0	528	45	0
2	P	284	0	158	13	0
2	R	284	0	158	16	0
2	T	284	0	158	10	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	U	284	0	158	6	0
2	W	284	0	158	7	0
2	Z	284	0	158	13	0
3	A	16	0	0	0	0
3	B	10	0	0	1	0
3	D	10	0	0	1	0
3	E	14	0	0	0	0
3	F	12	0	0	0	0
3	G	16	0	0	0	0
3	H	5	0	0	1	0
3	I	7	0	0	1	0
3	J	10	0	0	2	0
3	K	14	0	0	1	0
3	P	9	0	0	0	0
3	R	5	0	0	0	0
3	T	12	0	0	1	0
3	U	14	0	0	1	0
3	W	10	0	0	0	0
3	Z	4	0	0	0	0
All	All	7079	0	6251	303	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 23.

The worst 5 of 303 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:B:31:PRO:N	1:B:31:PRO:CD	1.71	1.32
1:K:32:ARG:HG3	1:K:32:ARG:HH11	1.14	1.08
2:Z:13:DC:H2''	2:Z:14:DA:H5'	1.37	1.07
1:E:14:LEU:O	1:E:54:LYS:NZ	1.98	0.95
1:I:62:ILE:HD11	1:I:64:LEU:HD21	1.47	0.95

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	65/69 (94%)	60 (92%)	5 (8%)	0	100	100
1	B	66/69 (96%)	58 (88%)	8 (12%)	0	100	100
1	D	65/69 (94%)	60 (92%)	5 (8%)	0	100	100
1	E	66/69 (96%)	59 (89%)	7 (11%)	0	100	100
1	F	66/69 (96%)	64 (97%)	2 (3%)	0	100	100
1	G	66/69 (96%)	63 (96%)	3 (4%)	0	100	100
1	H	62/69 (90%)	54 (87%)	8 (13%)	0	100	100
1	I	63/69 (91%)	60 (95%)	3 (5%)	0	100	100
1	J	65/69 (94%)	57 (88%)	8 (12%)	0	100	100
1	K	64/69 (93%)	62 (97%)	2 (3%)	0	100	100
All	All	648/690 (94%)	597 (92%)	51 (8%)	0	100	100

There are no Ramachandran outliers to report.

### 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	58/60 (97%)	50 (86%)	8 (14%)	4	6
1	B	59/60 (98%)	55 (93%)	4 (7%)	18	35
1	D	58/60 (97%)	52 (90%)	6 (10%)	8	15
1	E	59/60 (98%)	54 (92%)	5 (8%)	12	23

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	F	59/60 (98%)	54 (92%)	5 (8%)	12	23
1	G	60/60 (100%)	52 (87%)	8 (13%)	4	7
1	H	56/60 (93%)	50 (89%)	6 (11%)	8	13
1	I	57/60 (95%)	53 (93%)	4 (7%)	18	34
1	J	58/60 (97%)	53 (91%)	5 (9%)	12	23
1	K	58/60 (97%)	52 (90%)	6 (10%)	8	15
All	All	582/600 (97%)	525 (90%)	57 (10%)	9	17

5 of 57 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	31	GLU
1	F	3	SER
1	K	11	GLU
1	G	37	HIS
1	G	48	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 20 such sidechains are listed below:

Mol	Chain	Res	Type
1	G	25	GLN
1	F	35	ASN
1	J	6	ASN
1	E	29	ASN
1	E	35	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

### 6.1 Protein, DNA and RNA chains

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	67/69 (97%)	0.19	1 (1%) 74 69	39, 74, 100, 105	0
1	B	68/69 (98%)	0.08	1 (1%) 74 69	39, 67, 95, 102	0
1	D	67/69 (97%)	0.04	2 (2%) 51 43	30, 60, 85, 93	0
1	E	68/69 (98%)	0.15	3 (4%) 35 28	37, 66, 99, 115	0
1	F	68/69 (98%)	0.16	5 (7%) 15 11	31, 60, 91, 108	0
1	G	68/69 (98%)	0.22	2 (2%) 52 45	40, 66, 97, 119	0
1	H	64/69 (92%)	1.17	13 (20%) 1 0	70, 100, 116, 120	0
1	I	65/69 (94%)	1.09	16 (24%) 1 0	67, 95, 106, 112	0
1	J	67/69 (97%)	0.17	1 (1%) 74 69	54, 72, 101, 108	0
1	K	66/69 (95%)	0.17	2 (3%) 51 43	36, 69, 96, 104	0
2	P	14/14 (100%)	-0.59	0 100 100	38, 52, 72, 72	0
2	R	14/14 (100%)	-0.55	0 100 100	38, 51, 70, 70	0
2	T	14/14 (100%)	-0.45	0 100 100	40, 64, 75, 83	0
2	U	14/14 (100%)	-0.53	0 100 100	33, 41, 50, 52	0
2	W	14/14 (100%)	-0.63	0 100 100	36, 42, 46, 50	0
2	Z	14/14 (100%)	-0.46	0 100 100	46, 57, 75, 78	0
All	All	752/774 (97%)	0.24	46 (6%) 22 17	30, 70, 105, 120	0

The worst 5 of 46 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	I	63	HIS	9.0
1	I	64	LEU	7.7
1	H	43	VAL	6.3
1	H	46	LEU	6.0
1	H	59	ILE	5.6

## 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.