



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 10:28 pm GMT

PDB ID : 2HPG  
Title : The crystal structure of a thermophilic TRAP periplasmic binding protein  
Authors : Cuneo, M.J.; Hellinga, H.W.  
Deposited on : 2006-07-17  
Resolution : 1.90 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

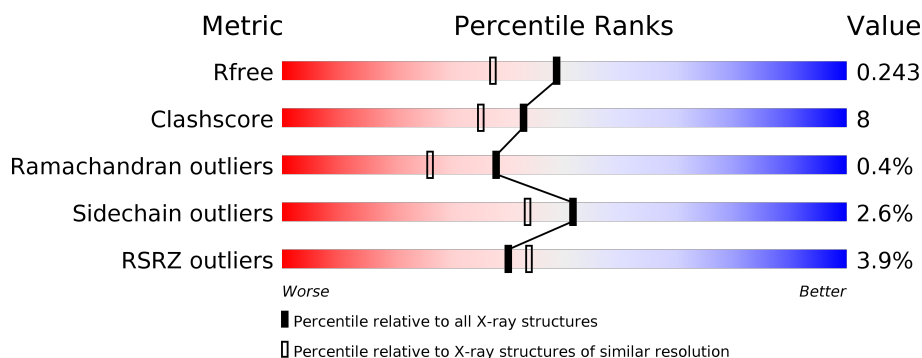
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.90 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	5047 (1.90-1.90)
Clashscore	112137	5731 (1.90-1.90)
Ramachandran outliers	110173	5669 (1.90-1.90)
Sidechain outliers	110143	5670 (1.90-1.90)
RSRZ outliers	101464	5100 (1.90-1.90)

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	327	<div> <div>3%</div> <div> <div></div> <div>85%</div> <div>11%</div> <div>..</div> </div> </div>
1	B	327	<div> <div>2%</div> <div> <div></div> <div>87%</div> <div>9%</div> <div>..</div> </div> </div>
1	C	327	<div> <div>5%</div> <div> <div></div> <div>79%</div> <div>15%</div> <div>..</div> </div> </div>
1	D	327	<div> <div>5%</div> <div> <div></div> <div>85%</div> <div>10%</div> <div>..</div> </div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 10882 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 ABC transporter, periplasmic substrate-binding protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	318	Total	C	N	O	S	Se	0	11	0
			2616	1693	427	480	1	15			
1	B	316	Total	C	N	O	S	Se	0	5	0
			2560	1654	420	474	1	11			
1	C	313	Total	C	N	O	S	Se	0	15	0
			2600	1684	421	481	1	13			
1	D	316	Total	C	N	O	S	Se	0	8	0
			2577	1672	423	471	1	10			

There are 68 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	16	MSE	-	INITIATING METHIONINE	UNP Q9WYF8
A	79	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
A	94	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
A	102	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
A	111	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
A	132	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
A	234	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
A	268	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
A	280	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
A	296	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
A	310	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
A	337	HIS	-	EXPRESSION TAG	UNP Q9WYF8
A	338	HIS	-	EXPRESSION TAG	UNP Q9WYF8
A	339	HIS	-	EXPRESSION TAG	UNP Q9WYF8
A	340	HIS	-	EXPRESSION TAG	UNP Q9WYF8
A	341	HIS	-	EXPRESSION TAG	UNP Q9WYF8
A	342	HIS	-	EXPRESSION TAG	UNP Q9WYF8
B	16	MSE	-	INITIATING METHIONINE	UNP Q9WYF8
B	79	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
B	94	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
B	102	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	111	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
B	132	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
B	234	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
B	268	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
B	280	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
B	296	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
B	310	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
B	337	HIS	-	EXPRESSION TAG	UNP Q9WYF8
B	338	HIS	-	EXPRESSION TAG	UNP Q9WYF8
B	339	HIS	-	EXPRESSION TAG	UNP Q9WYF8
B	340	HIS	-	EXPRESSION TAG	UNP Q9WYF8
B	341	HIS	-	EXPRESSION TAG	UNP Q9WYF8
B	342	HIS	-	EXPRESSION TAG	UNP Q9WYF8
C	16	MSE	-	INITIATING METHIONINE	UNP Q9WYF8
C	79	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
C	94	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
C	102	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
C	111	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
C	132	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
C	234	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
C	268	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
C	280	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
C	296	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
C	310	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
C	337	HIS	-	EXPRESSION TAG	UNP Q9WYF8
C	338	HIS	-	EXPRESSION TAG	UNP Q9WYF8
C	339	HIS	-	EXPRESSION TAG	UNP Q9WYF8
C	340	HIS	-	EXPRESSION TAG	UNP Q9WYF8
C	341	HIS	-	EXPRESSION TAG	UNP Q9WYF8
C	342	HIS	-	EXPRESSION TAG	UNP Q9WYF8
D	16	MSE	-	INITIATING METHIONINE	UNP Q9WYF8
D	79	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
D	94	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
D	102	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
D	111	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
D	132	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
D	234	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
D	268	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
D	280	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
D	296	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
D	310	MSE	MET	MODIFIED RESIDUE	UNP Q9WYF8
D	337	HIS	-	EXPRESSION TAG	UNP Q9WYF8

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
D	338	HIS	-	EXPRESSION TAG	UNP Q9WYF8
D	339	HIS	-	EXPRESSION TAG	UNP Q9WYF8
D	340	HIS	-	EXPRESSION TAG	UNP Q9WYF8
D	341	HIS	-	EXPRESSION TAG	UNP Q9WYF8
D	342	HIS	-	EXPRESSION TAG	UNP Q9WYF8

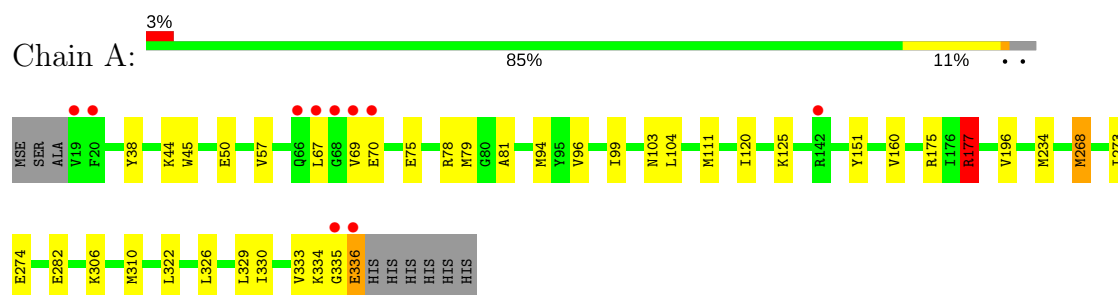
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	180	Total	O	0	0
			180	180		
2	B	120	Total	O	0	0
			120	120		
2	C	150	Total	O	0	0
			150	150		
2	D	79	Total	O	0	0
			79	79		

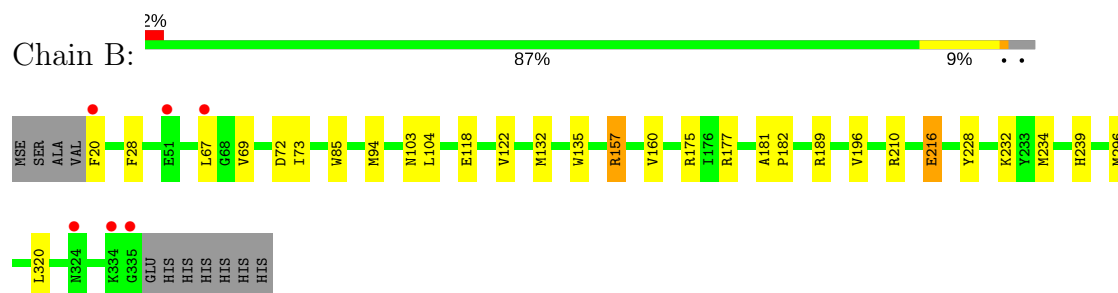
### 3 Residue-property plots [i](#)

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.

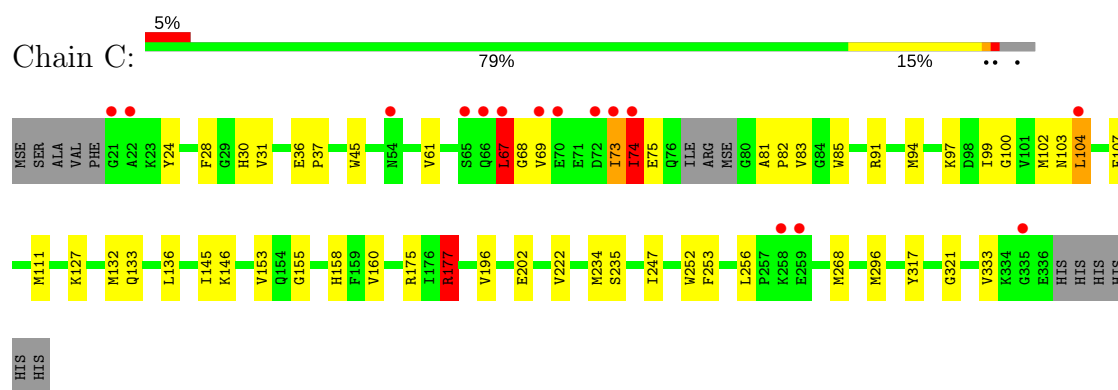
- Molecule 1: ABC transporter, periplasmic substrate-binding protein



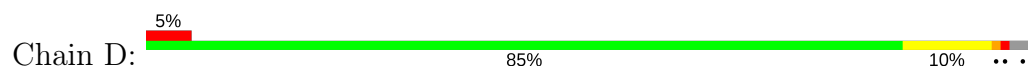
- Molecule 1: ABC transporter, periplasmic substrate-binding protein

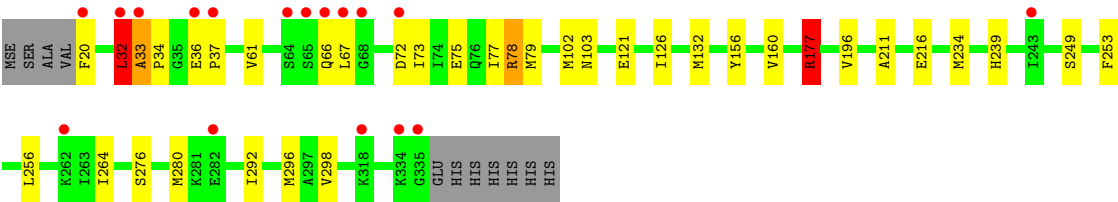


- Molecule 1: ABC transporter, periplasmic substrate-binding protein



- Molecule 1: ABC transporter, periplasmic substrate-binding protein





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 65 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	119.26Å 119.26Å 429.10Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	30.00 – 1.90 29.82 – 1.90	Depositor EDS
% Data completeness (in resolution range)	100.0 (30.00-1.90) 100.0 (29.82-1.90)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	0.09	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.23 (at 1.91Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.213 , 0.239 0.218 , 0.243	Depositor DCC
$R_{free}$ test set	7152 reflections (5.29%)	DCC
Wilson B-factor (Å <sup>2</sup> )	23.6	Xtriage
Anisotropy	0.424	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.40 , 49.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	10882	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	28.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.30% 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

### 5.1 Standard geometry

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	0.56	0/2685	0.66	2/3600 (0.1%)
1	B	0.56	1/2618 (0.0%)	0.66	4/3514 (0.1%)
1	C	0.54	0/2681	0.64	3/3600 (0.1%)
1	D	0.49	0/2644	0.62	2/3549 (0.1%)
All	All	0.54	1/10628 (0.0%)	0.65	11/14263 (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
1	C	0	1
1	D	0	1
All	All	0	2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	216	GLU	CB-CG	-6.44	1.40	1.52

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	177	ARG	NE-CZ-NH1	11.12	125.86	120.30
1	B	157	ARG	NE-CZ-NH2	-10.34	115.13	120.30
1	A	177	ARG	NE-CZ-NH1	9.13	124.87	120.30
1	C	177	ARG	NE-CZ-NH2	-9.06	115.77	120.30
1	D	177	ARG	NE-CZ-NH1	8.13	124.37	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	74	ILE	Peptide
1	D	32	LEU	Peptide

## 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	2616	0	2651	45	0
1	B	2560	0	2576	40	0
1	C	2600	0	2620	60	0
1	D	2577	0	2623	39	0
2	A	180	0	0	0	0
2	B	120	0	0	0	0
2	C	150	0	0	2	0
2	D	79	0	0	2	0
All	All	10882	0	10470	162	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 8.

The worst 5 of 162 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:228:TYR:HB3	1:B:296:MSE:CE	1.68	1.23
1:B:228:TYR:HB3	1:B:296:MSE:HE2	1.24	1.11
1:C:73:ILE:HG22	1:C:73:ILE:O	1.56	1.03
1:B:94[B]:MSE:CE	1:B:181:ALA:HB1	1.90	1.02
1:B:228:TYR:CB	1:B:296:MSE:HE2	1.90	1.00

There are no symmetry-related clashes.

## 5.3 Torsion angles ⓘ

### 5.3.1 Protein backbone ⓘ

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	327/327 (100%)	318 (97%)	8 (2%)	1 (0%)	44	34
1	B	319/327 (98%)	311 (98%)	8 (2%)	0	100	100
1	C	324/327 (99%)	312 (96%)	9 (3%)	3 (1%)	20	8
1	D	322/327 (98%)	316 (98%)	5 (2%)	1 (0%)	44	34
All	All	1292/1308 (99%)	1257 (97%)	30 (2%)	5 (0%)	38	26

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	33	ALA
1	A	334	LYS
1	C	74	ILE
1	C	67	LEU
1	C	73	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	
1	A	279/265 (105%)	269 (96%)	10 (4%)	40	29
1	B	271/265 (102%)	265 (98%)	6 (2%)	57	51
1	C	278/265 (105%)	270 (97%)	8 (3%)	48	39
1	D	274/265 (103%)	267 (97%)	7 (3%)	51	43
All	All	1102/1060 (104%)	1071 (97%)	31 (3%)	51	40

5 of 31 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	210	ARG
1	C	91	ARG

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	D	103	ASN
1	B	232	LYS
1	C	103	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	261	GLN
1	D	76	GLN
1	B	154	GLN
1	A	103	ASN
1	C	328	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	308/327 (94%)	-0.00	10 (3%) 48 51	13, 23, 43, 54	0
1	B	306/327 (93%)	-0.03	6 (1%) 65 69	14, 26, 37, 45	0
1	C	304/327 (92%)	0.13	15 (4%) 30 34	13, 24, 48, 68	0
1	D	306/327 (93%)	0.33	17 (5%) 25 28	24, 35, 45, 56	0
All	All	1224/1308 (93%)	0.11	48 (3%) 40 44	13, 28, 44, 68	0

The worst 5 of 48 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	67	LEU	10.0
1	C	21	GLY	7.3
1	A	19	VAL	7.3
1	C	69	VAL	7.0
1	C	66	GLN	6.7

### 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.