



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 17, 2016 – 11:44 PM EDT

PDB ID : 5CM5  
Title : Structure of Hydroxyethylthiazole Kinase ThiM from *Staphylococcus aureus*  
Authors : Drebes, J.; Kuenz, M.; Eberle, R.J.; Oberthuer, D.; Cang, H.; Wrenger, C.; Betzel, C.  
Deposited on : 2015-07-16  
Resolution : 2.09 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 : unknown  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20027107  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0122  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20027107

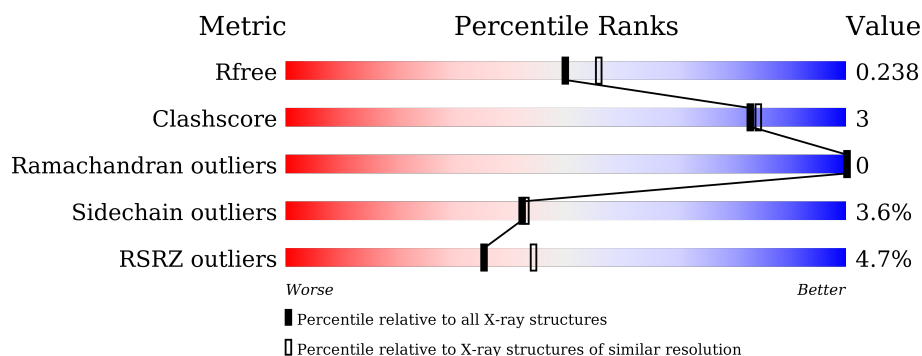
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.09 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	3939 (2.10-2.10)
Clashscore	102246	4460 (2.10-2.10)
Ramachandran outliers	100387	4413 (2.10-2.10)
Sidechain outliers	100360	4414 (2.10-2.10)
RSRZ outliers	91569	3948 (2.10-2.10)

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	277	<div> <div>4%</div> <div>85% 9% • 5%</div> </div>
1	B	277	<div> <div>3%</div> <div>83% 7% • 8%</div> </div>
1	C	277	<div> <div>3%</div> <div>83% 8% • 7%</div> </div>
1	D	277	<div> <div>4%</div> <div>83% 6% • 8%</div> </div>
1	E	277	<div> <div>4%</div> <div>82% 7% • 10%</div> </div>
1	F	277	<div> <div>9%</div> <div>81% 8% • 10%</div> </div>

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 12026 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 Hydroxyethylthiazole kinase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	262	Total	C	N	O	S	0	1	0
			1965	1245	323	390	7			
1	B	254	Total	C	N	O	S	0	2	0
			1929	1231	313	379	6			
1	C	258	Total	C	N	O	S	0	0	0
			1933	1230	317	379	7			
1	D	254	Total	C	N	O	S	0	0	0
			1909	1215	312	376	6			
1	E	249	Total	C	N	O	S	0	1	0
			1874	1197	307	364	6			
1	F	248	Total	C	N	O	S	0	1	0
			1864	1189	307	362	6			

There are 84 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	264	GLU	-	expression tag	UNP Q6GEY3
A	265	ASN	-	expression tag	UNP Q6GEY3
A	266	LEU	-	expression tag	UNP Q6GEY3
A	267	TYR	-	expression tag	UNP Q6GEY3
A	268	PHE	-	expression tag	UNP Q6GEY3
A	269	GLN	-	expression tag	UNP Q6GEY3
A	270	SER	-	expression tag	UNP Q6GEY3
A	271	GLY	-	expression tag	UNP Q6GEY3
A	272	HIS	-	expression tag	UNP Q6GEY3
A	273	HIS	-	expression tag	UNP Q6GEY3
A	274	HIS	-	expression tag	UNP Q6GEY3
A	275	HIS	-	expression tag	UNP Q6GEY3
A	276	HIS	-	expression tag	UNP Q6GEY3
A	277	HIS	-	expression tag	UNP Q6GEY3
B	264	GLU	-	expression tag	UNP Q6GEY3
B	265	ASN	-	expression tag	UNP Q6GEY3
B	266	LEU	-	expression tag	UNP Q6GEY3

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	267	TYR	-	expression tag	UNP Q6GEY3
B	268	PHE	-	expression tag	UNP Q6GEY3
B	269	GLN	-	expression tag	UNP Q6GEY3
B	270	SER	-	expression tag	UNP Q6GEY3
B	271	GLY	-	expression tag	UNP Q6GEY3
B	272	HIS	-	expression tag	UNP Q6GEY3
B	273	HIS	-	expression tag	UNP Q6GEY3
B	274	HIS	-	expression tag	UNP Q6GEY3
B	275	HIS	-	expression tag	UNP Q6GEY3
B	276	HIS	-	expression tag	UNP Q6GEY3
B	277	HIS	-	expression tag	UNP Q6GEY3
C	264	GLU	-	expression tag	UNP Q6GEY3
C	265	ASN	-	expression tag	UNP Q6GEY3
C	266	LEU	-	expression tag	UNP Q6GEY3
C	267	TYR	-	expression tag	UNP Q6GEY3
C	268	PHE	-	expression tag	UNP Q6GEY3
C	269	GLN	-	expression tag	UNP Q6GEY3
C	270	SER	-	expression tag	UNP Q6GEY3
C	271	GLY	-	expression tag	UNP Q6GEY3
C	272	HIS	-	expression tag	UNP Q6GEY3
C	273	HIS	-	expression tag	UNP Q6GEY3
C	274	HIS	-	expression tag	UNP Q6GEY3
C	275	HIS	-	expression tag	UNP Q6GEY3
C	276	HIS	-	expression tag	UNP Q6GEY3
C	277	HIS	-	expression tag	UNP Q6GEY3
D	264	GLU	-	expression tag	UNP Q6GEY3
D	265	ASN	-	expression tag	UNP Q6GEY3
D	266	LEU	-	expression tag	UNP Q6GEY3
D	267	TYR	-	expression tag	UNP Q6GEY3
D	268	PHE	-	expression tag	UNP Q6GEY3
D	269	GLN	-	expression tag	UNP Q6GEY3
D	270	SER	-	expression tag	UNP Q6GEY3
D	271	GLY	-	expression tag	UNP Q6GEY3
D	272	HIS	-	expression tag	UNP Q6GEY3
D	273	HIS	-	expression tag	UNP Q6GEY3
D	274	HIS	-	expression tag	UNP Q6GEY3
D	275	HIS	-	expression tag	UNP Q6GEY3
D	276	HIS	-	expression tag	UNP Q6GEY3
D	277	HIS	-	expression tag	UNP Q6GEY3
E	264	GLU	-	expression tag	UNP Q6GEY3
E	265	ASN	-	expression tag	UNP Q6GEY3
E	266	LEU	-	expression tag	UNP Q6GEY3

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
E	267	TYR	-	expression tag	UNP Q6GEY3
E	268	PHE	-	expression tag	UNP Q6GEY3
E	269	GLN	-	expression tag	UNP Q6GEY3
E	270	SER	-	expression tag	UNP Q6GEY3
E	271	GLY	-	expression tag	UNP Q6GEY3
E	272	HIS	-	expression tag	UNP Q6GEY3
E	273	HIS	-	expression tag	UNP Q6GEY3
E	274	HIS	-	expression tag	UNP Q6GEY3
E	275	HIS	-	expression tag	UNP Q6GEY3
E	276	HIS	-	expression tag	UNP Q6GEY3
E	277	HIS	-	expression tag	UNP Q6GEY3
F	264	GLU	-	expression tag	UNP Q6GEY3
F	265	ASN	-	expression tag	UNP Q6GEY3
F	266	LEU	-	expression tag	UNP Q6GEY3
F	267	TYR	-	expression tag	UNP Q6GEY3
F	268	PHE	-	expression tag	UNP Q6GEY3
F	269	GLN	-	expression tag	UNP Q6GEY3
F	270	SER	-	expression tag	UNP Q6GEY3
F	271	GLY	-	expression tag	UNP Q6GEY3
F	272	HIS	-	expression tag	UNP Q6GEY3
F	273	HIS	-	expression tag	UNP Q6GEY3
F	274	HIS	-	expression tag	UNP Q6GEY3
F	275	HIS	-	expression tag	UNP Q6GEY3
F	276	HIS	-	expression tag	UNP Q6GEY3
F	277	HIS	-	expression tag	UNP Q6GEY3

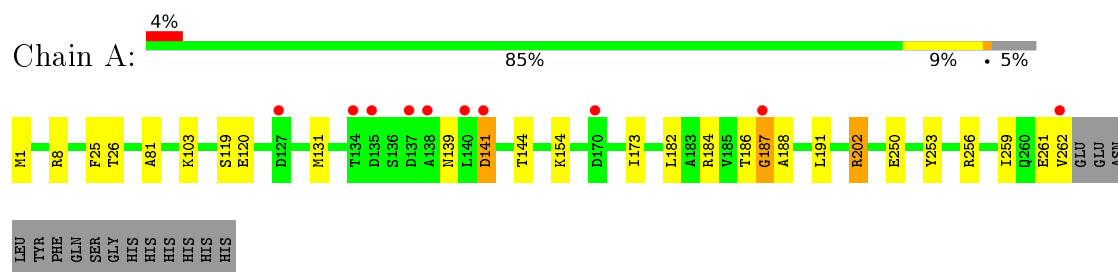
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	115	Total O 115 115	0	0
2	B	114	Total O 114 114	0	0
2	C	136	Total O 136 136	0	0
2	D	69	Total O 69 69	0	0
2	E	77	Total O 77 77	0	0
2	F	41	Total O 41 41	0	0

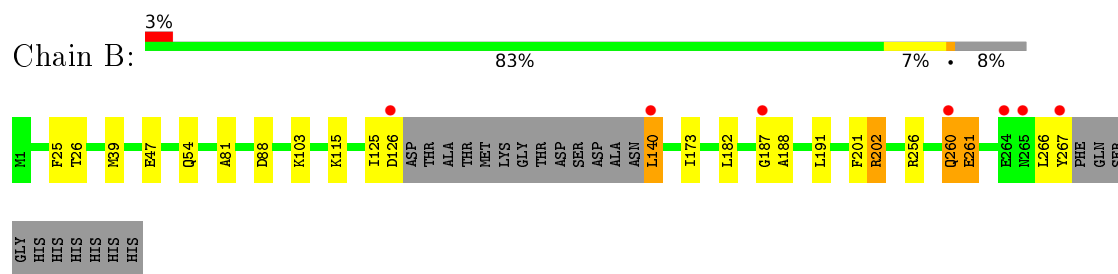
### 3 Residue-property plots [i](#)

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

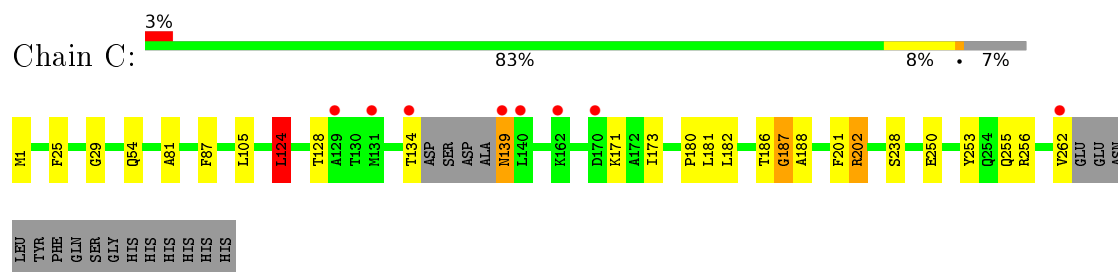
- Molecule 1: Hydroxyethylthiazole kinase



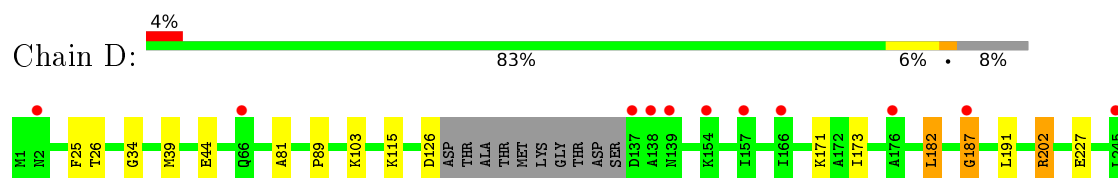
- Molecule 1: Hydroxyethylthiazole kinase



- Molecule 1: Hydroxyethylthiazole kinase



- Molecule 1: Hydroxyethylthiazole kinase





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	62.56Å 103.52Å 126.18Å 90.00° 99.48° 90.00°	Depositor
Resolution (Å)	20.00 – 2.09 19.99 – 2.09	Depositor EDS
% Data completeness (in resolution range)	99.8 (20.00-2.09) 99.9 (19.99-2.09)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.09 (at 2.09Å)	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
R, $R_{free}$	0.203 , 0.230 0.210 , 0.238	Depositor DCC
$R_{free}$ test set	4693 reflections (5.27%)	DCC
Wilson B-factor (Å <sup>2</sup> )	27.8	Xtriage
Anisotropy	0.617	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 36.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Outliers	0 of 93703 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	12026	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.27% 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 ⓘ

### 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	0.86	0/1990	0.85	3/2709 (0.1%)
1	B	0.89	5/1957 (0.3%)	0.82	4/2662 (0.2%)
1	C	0.87	2/1957 (0.1%)	0.82	3/2661 (0.1%)
1	D	0.72	1/1933 (0.1%)	0.82	5/2629 (0.2%)
1	E	0.63	0/1901	0.74	2/2586 (0.1%)
1	F	0.62	0/1888	0.75	3/2568 (0.1%)
All	All	0.77	8/11626 (0.1%)	0.80	20/15815 (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	A	0	2
1	B	0	1
1	C	0	1
1	D	0	1
All	All	0	5

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	260	GLN	N-CA	7.34	1.61	1.46
1	B	47[A]	GLU	CD-OE1	7.07	1.33	1.25
1	B	47[B]	GLU	CD-OE1	7.07	1.33	1.25
1	C	29	GLY	C-O	6.26	1.33	1.23
1	B	260	GLN	N-CA	5.98	1.58	1.46

The worst 5 of 20 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	182	LEU	CA-CB-CG	9.07	136.15	115.30

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	F	259	ILE	CB-CA-C	-7.56	96.48	111.60
1	B	256	ARG	NE-CZ-NH2	-6.31	117.15	120.30
1	B	261	GLU	N-CA-C	6.29	127.99	111.00
1	D	256	ARG	NE-CZ-NH2	-6.27	117.16	120.30

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	139	ASN	Peptide
1	A	187	GLY	Peptide
1	B	140	LEU	Peptide
1	C	187	GLY	Peptide
1	D	187	GLY	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	1965	0	1994	18	0
1	B	1929	0	1965	8	0
1	C	1933	0	1978	13	0
1	D	1909	0	1944	17	0
1	E	1874	0	1923	9	0
1	F	1864	0	1903	11	0
2	A	115	0	0	1	0
2	B	114	0	0	0	0
2	C	136	0	0	2	0
2	D	69	0	0	3	0
2	E	77	0	0	1	0
2	F	41	0	0	1	0
All	All	12026	0	11707	71	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 3.

The worst 5 of 71 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:D:259:ILE:C	1:D:260:GLN:OE1	2.04	0.96
1:A:8:ARG:HB2	1:D:261:GLU:OE1	1.66	0.94
1:D:260:GLN:NE2	2:D:301:HOH:O	2.13	0.80
1:D:259:ILE:O	1:D:260:GLN:OE1	2.10	0.69
1:C:139:ASN:OD1	1:C:139:ASN:N	2.26	0.68

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	261/277 (94%)	259 (99%)	2 (1%)	0	100	100
1	B	252/277 (91%)	246 (98%)	6 (2%)	0	100	100
1	C	254/277 (92%)	251 (99%)	3 (1%)	0	100	100
1	D	250/277 (90%)	245 (98%)	5 (2%)	0	100	100
1	E	246/277 (89%)	243 (99%)	3 (1%)	0	100	100
1	F	245/277 (88%)	241 (98%)	4 (2%)	0	100	100
All	All	1508/1662 (91%)	1485 (98%)	23 (2%)	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	209/223 (94%)	204 (98%)	5 (2%)	57	61
1	B	206/223 (92%)	197 (96%)	9 (4%)	35	33
1	C	206/223 (92%)	198 (96%)	8 (4%)	39	39
1	D	203/223 (91%)	197 (97%)	6 (3%)	48	51
1	E	200/223 (90%)	191 (96%)	9 (4%)	34	32
1	F	197/223 (88%)	190 (96%)	7 (4%)	42	43
All	All	1221/1338 (91%)	1177 (96%)	44 (4%)	42	43

5 of 44 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	181	LEU
1	D	173	ILE
1	F	154	LYS
1	C	202	ARG
1	D	39	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	100	GLN
1	E	54	GLN
1	F	6	ASN
1	C	219	ASN
1	F	5	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

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	262/277 (94%)	0.17	10 (3%) 44 53	16, 30, 54, 80	3 (1%)
1	B	254/277 (91%)	0.04	7 (2%) 56 64	18, 29, 55, 86	3 (1%)
1	C	258/277 (93%)	0.02	8 (3%) 52 61	17, 30, 55, 73	2 (0%)
1	D	254/277 (91%)	0.22	12 (4%) 35 44	25, 39, 60, 87	1 (0%)
1	E	249/277 (89%)	0.24	11 (4%) 38 47	24, 41, 70, 96	4 (1%)
1	F	248/277 (89%)	0.54	24 (9%) 10 14	26, 49, 74, 105	2 (0%)
All	All	1525/1662 (91%)	0.20	72 (4%) 35 44	16, 36, 67, 105	15 (0%)

The worst 5 of 72 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	140	LEU	10.2
1	D	138	ALA	6.0
1	E	178	GLY	5.5
1	C	139	ASN	4.9
1	E	187	GLY	4.7

### 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.