



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 2, 2021 – 05:34 PM EDT

PDB ID : 2E2K  
Title : Helicobacter pylori formamidase AmiF contains a fine-tuned cysteine-glutamate-lysine catalytic triad  
Authors : Wang, W.C.; Hung, C.L.  
Deposited on : 2006-11-14  
Resolution : 2.50 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

<https://www.wwpdb.org/validation/2017/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)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

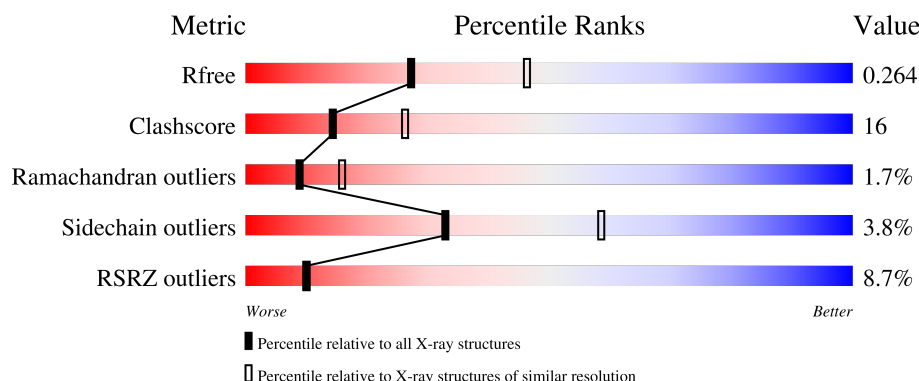
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.50 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))
$R_{free}$	130704	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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 of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria respectively. 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	334	<div> <div>8%</div> <div> <div></div> <div>75%</div> <div>16%</div> <div>• 6%</div> </div> </div>
1	B	334	<div> <div>7%</div> <div> <div></div> <div>74%</div> <div>18%</div> <div>• 6%</div> </div> </div>
1	C	334	<div> <div>7%</div> <div> <div></div> <div>74%</div> <div>17%</div> <div>• 6%</div> </div> </div>
1	D	334	<div> <div>7%</div> <div> <div></div> <div>73%</div> <div>18%</div> <div>• 7%</div> </div> </div>
1	E	334	<div> <div>9%</div> <div> <div></div> <div>74%</div> <div>17%</div> <div>• 6%</div> </div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	334	<div><div></div><div>10%</div><div>77%</div><div>14%</div><div>• 6%</div></div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 15744 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 Formamidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	313	Total	C	N	O	S	0	0	0
			2482	1597	416	458	11			
1	B	314	Total	C	N	O	S	0	0	0
			2485	1598	417	459	11			
1	C	314	Total	C	N	O	S	0	0	0
			2489	1600	417	461	11			
1	D	311	Total	C	N	O	S	0	0	0
			2458	1581	412	454	11			
1	E	315	Total	C	N	O	S	0	0	0
			2493	1602	419	461	11			
1	F	314	Total	C	N	O	S	0	0	0
			2487	1601	415	460	11			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	166	SER	CYS	engineered mutation	UNP O25836
B	166	SER	CYS	engineered mutation	UNP O25836
C	166	SER	CYS	engineered mutation	UNP O25836
D	166	SER	CYS	engineered mutation	UNP O25836
E	166	SER	CYS	engineered mutation	UNP O25836
F	166	SER	CYS	engineered mutation	UNP O25836

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	148	Total	O	0	0
			148	148		
2	B	151	Total	O	0	0
			151	151		
2	C	133	Total	O	0	0
			133	133		

*Continued on next page...*

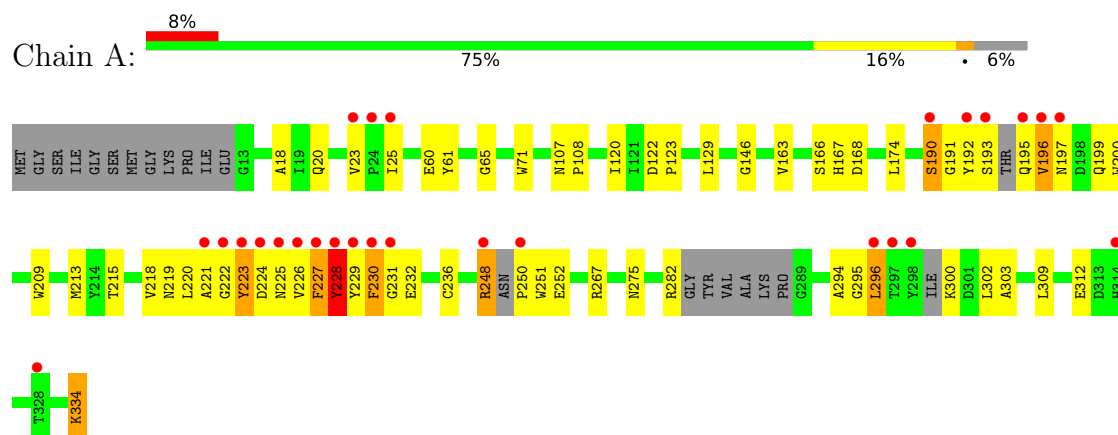
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	D	159	Total 159	O 159	0	0
2	E	138	Total 138	O 138	0	0
2	F	121	Total 121	O 121	0	0

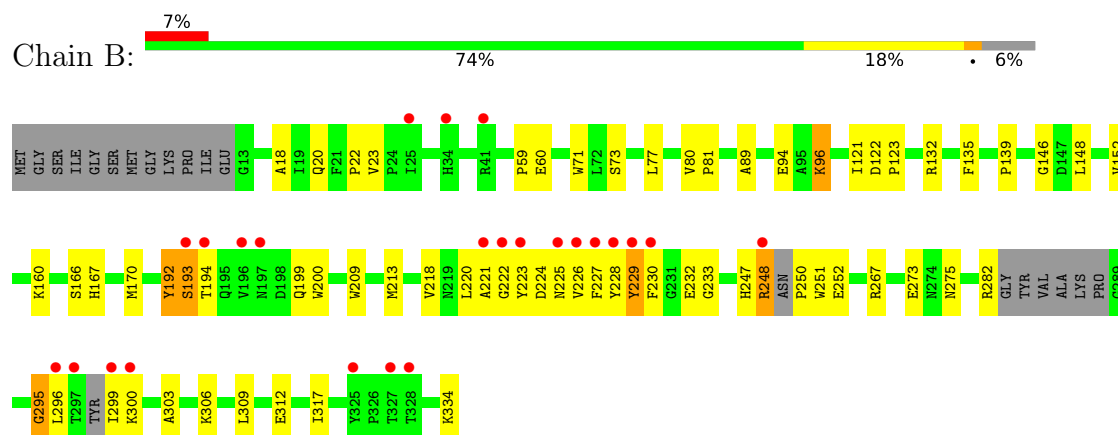
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

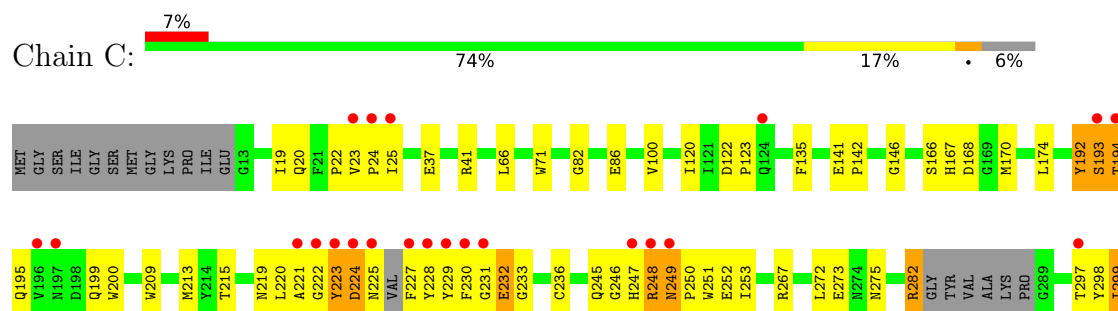
#### • Molecule 1: Formamidase



#### • Molecule 1: Formamidase

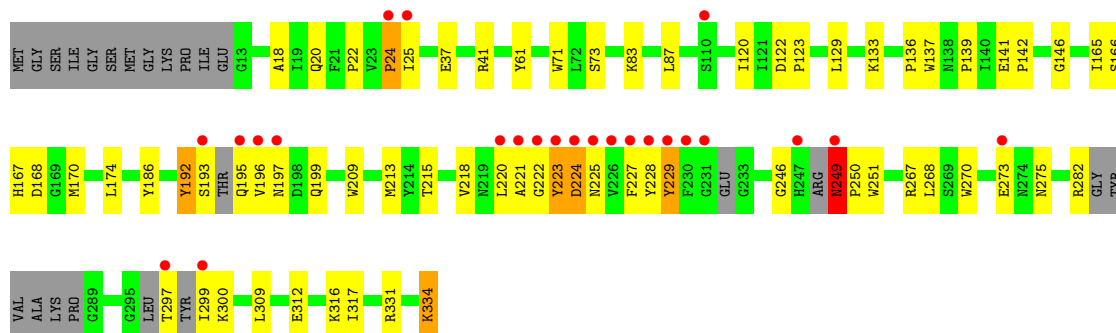
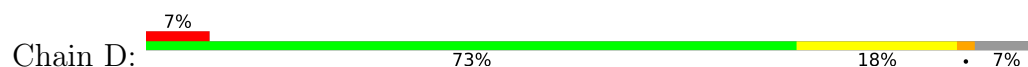


#### • Molecule 1: Formamidase

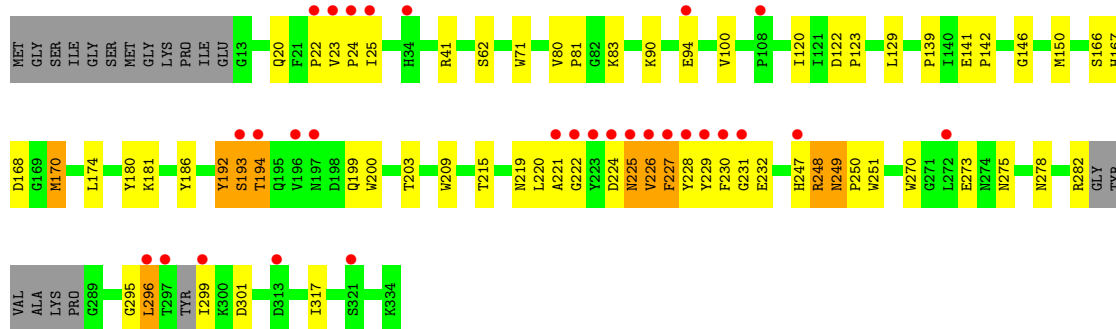
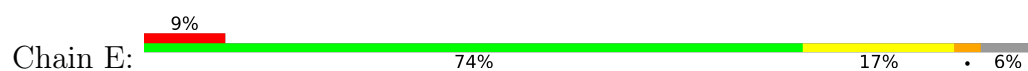




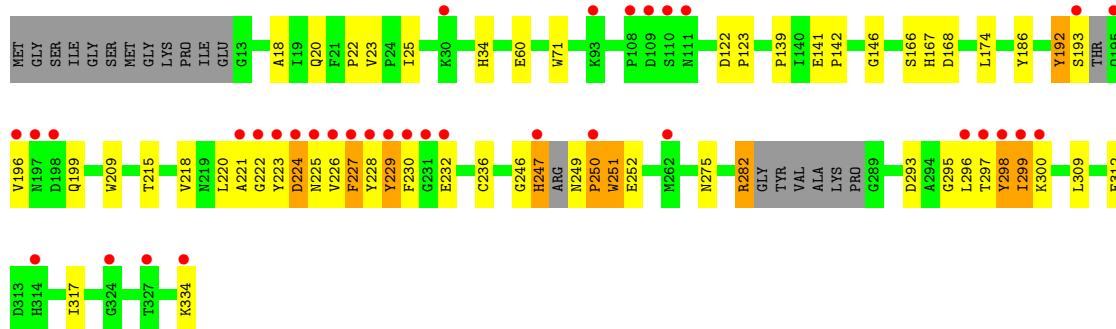
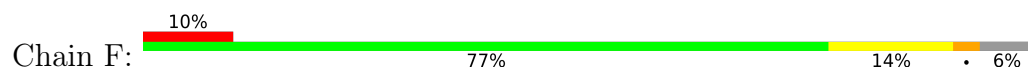
● Molecule 1: Formamidase



● Molecule 1: Formamidase



● Molecule 1: Formamidase



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	117.72Å 130.53Å 144.59Å 90.00° 99.44° 90.00°	Depositor
Resolution (Å)	30.00 – 2.50 26.90 – 2.50	Depositor EDS
% Data completeness (in resolution range)	99.8 (30.00-2.50) 99.8 (26.90-2.50)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	13.24 (at 2.50Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.256 , 0.294 0.256 , 0.264	Depositor DCC
$R_{free}$ test set	3760 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	13.2	Xtriage
Anisotropy	0.292	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 31.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.86	EDS
Total number of atoms	15744	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	9.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.95% 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	0.26	0/2550	0.46	0/3459
1	B	0.27	0/2553	0.47	0/3465
1	C	0.26	0/2558	0.47	0/3474
1	D	0.27	0/2523	0.46	0/3421
1	E	0.27	0/2562	0.46	0/3480
1	F	0.26	0/2556	0.47	0/3471
All	All	0.27	0/15302	0.46	0/20770

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	D	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	249	ASN	Peptide

### 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	2482	0	2407	111	0
1	B	2485	0	2417	84	0
1	C	2489	0	2409	101	0
1	D	2458	0	2382	83	0
1	E	2493	0	2423	92	0
1	F	2487	0	2411	75	0
2	A	148	0	0	6	0
2	B	151	0	0	1	0
2	C	133	0	0	6	0
2	D	159	0	0	4	0
2	E	138	0	0	3	0
2	F	121	0	0	0	0
All	All	15744	0	14449	455	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 16.

The worst 5 of 455 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:F:192:TYR:HB3	1:F:193:SER:CA	1.69	1.23
1:E:248:ARG:H	1:E:248:ARG:HD3	1.06	1.20
1:A:192:TYR:CB	1:A:193:SER:HA	1.73	1.18
1:D:192:TYR:HB3	1:D:193:SER:CA	1.76	1.14
1:C:192:TYR:HB3	1:C:193:SER:CA	1.78	1.13

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	303/334 (91%)	282 (93%)	19 (6%)	2 (1%)	22 39

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	306/334 (92%)	282 (92%)	18 (6%)	6 (2%)	7	12
1	C	306/334 (92%)	287 (94%)	12 (4%)	7 (2%)	6	10
1	D	298/334 (89%)	278 (93%)	16 (5%)	4 (1%)	12	21
1	E	309/334 (92%)	281 (91%)	22 (7%)	6 (2%)	8	13
1	F	306/334 (92%)	278 (91%)	22 (7%)	6 (2%)	7	12
All	All	1828/2004 (91%)	1688 (92%)	109 (6%)	31 (2%)	9	16

5 of 31 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	194	THR
1	C	194	THR
1	E	194	THR
1	E	226	VAL
1	F	250	PRO

### 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	261/277 (94%)	247 (95%)	14 (5%)	22	42
1	B	262/277 (95%)	252 (96%)	10 (4%)	33	58
1	C	262/277 (95%)	254 (97%)	8 (3%)	40	67
1	D	259/277 (94%)	252 (97%)	7 (3%)	44	71
1	E	263/277 (95%)	254 (97%)	9 (3%)	37	63
1	F	262/277 (95%)	251 (96%)	11 (4%)	30	54
All	All	1569/1662 (94%)	1510 (96%)	59 (4%)	33	58

5 of 59 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	282	ARG

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	F	247	HIS
1	D	229	TYR
1	F	232	GLU
1	F	71	TRP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 42 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	281	HIS
1	E	275	ASN
1	E	20	GLN
1	E	210	HIS
1	F	20	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 monosaccharides 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	313/334 (93%)	0.47	27 (8%)	10 10	2, 8, 20, 28	0
1	B	314/334 (94%)	0.41	24 (7%)	13 14	3, 8, 18, 25	0
1	C	314/334 (94%)	0.35	24 (7%)	13 14	2, 7, 15, 24	0
1	D	311/334 (93%)	0.40	24 (7%)	13 13	2, 7, 17, 28	0
1	E	315/334 (94%)	0.59	29 (9%)	9 9	3, 10, 21, 29	0
1	F	314/334 (94%)	0.59	35 (11%)	5 5	4, 10, 21, 26	0
All	All	1881/2004 (93%)	0.47	163 (8%)	10 10	2, 8, 20, 29	0

The worst 5 of 163 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	222	GLY	13.2
1	A	222	GLY	12.3
1	B	222	GLY	11.7
1	E	228	TYR	10.8
1	F	228	TYR	10.5

### 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 monosaccharides 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.