



# wwPDB X-ray Structure Validation Summary Report

Mar 13, 2024 – 04:08 PM JST

PDB ID : 8HZ4  
Title : The tetrameric structure of biotin carboxylase from *Chloroflexus aurantiacus* in complex with bicarbonate  
Authors : Shen, J.; Wu, W.; Xu, X.  
Deposited on : 2023-01-08  
Resolution : 3.20 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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 : 2.36  
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) : 2.36

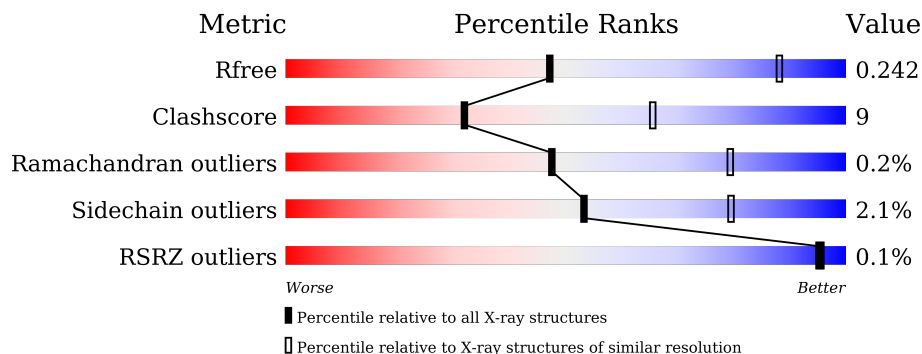
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.20 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.



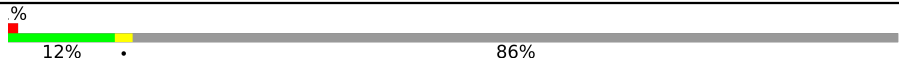

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	1133 (3.20-3.20)
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)
RSRZ outliers	127900	1095 (3.20-3.20)

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	469	
1	B	469	
1	C	469	
1	D	469	
2	E	137	
2	F	137	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	G	137	 12% 86%
2	H	137	 11% 87%

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 14386 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 Biotin carboxylase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	456	Total 3509	C 2222	N 631	O 643	S 13	0	0	0
1	B	453	Total 3484	C 2210	N 623	O 638	S 13	0	0	0
1	C	434	Total 3343	C 2117	N 603	O 610	S 13	0	0	0
1	D	439	Total 3392	C 2150	N 607	O 622	S 13	0	0	0

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-9	MET	-	initiating methionine	UNP A9W9X0
A	-8	GLY	-	expression tag	UNP A9W9X0
A	-7	SER	-	expression tag	UNP A9W9X0
A	-6	SER	-	expression tag	UNP A9W9X0
A	-5	HIS	-	expression tag	UNP A9W9X0
A	-4	HIS	-	expression tag	UNP A9W9X0
A	-3	HIS	-	expression tag	UNP A9W9X0
A	-2	HIS	-	expression tag	UNP A9W9X0
A	-1	HIS	-	expression tag	UNP A9W9X0
A	0	HIS	-	expression tag	UNP A9W9X0
B	-9	MET	-	initiating methionine	UNP A9W9X0
B	-8	GLY	-	expression tag	UNP A9W9X0
B	-7	SER	-	expression tag	UNP A9W9X0
B	-6	SER	-	expression tag	UNP A9W9X0
B	-5	HIS	-	expression tag	UNP A9W9X0
B	-4	HIS	-	expression tag	UNP A9W9X0
B	-3	HIS	-	expression tag	UNP A9W9X0
B	-2	HIS	-	expression tag	UNP A9W9X0
B	-1	HIS	-	expression tag	UNP A9W9X0
B	0	HIS	-	expression tag	UNP A9W9X0
C	-9	MET	-	initiating methionine	UNP A9W9X0

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	-8	GLY	-	expression tag	UNP A9W9X0
C	-7	SER	-	expression tag	UNP A9W9X0
C	-6	SER	-	expression tag	UNP A9W9X0
C	-5	HIS	-	expression tag	UNP A9W9X0
C	-4	HIS	-	expression tag	UNP A9W9X0
C	-3	HIS	-	expression tag	UNP A9W9X0
C	-2	HIS	-	expression tag	UNP A9W9X0
C	-1	HIS	-	expression tag	UNP A9W9X0
C	0	HIS	-	expression tag	UNP A9W9X0
D	-9	MET	-	initiating methionine	UNP A9W9X0
D	-8	GLY	-	expression tag	UNP A9W9X0
D	-7	SER	-	expression tag	UNP A9W9X0
D	-6	SER	-	expression tag	UNP A9W9X0
D	-5	HIS	-	expression tag	UNP A9W9X0
D	-4	HIS	-	expression tag	UNP A9W9X0
D	-3	HIS	-	expression tag	UNP A9W9X0
D	-2	HIS	-	expression tag	UNP A9W9X0
D	-1	HIS	-	expression tag	UNP A9W9X0
D	0	HIS	-	expression tag	UNP A9W9X0

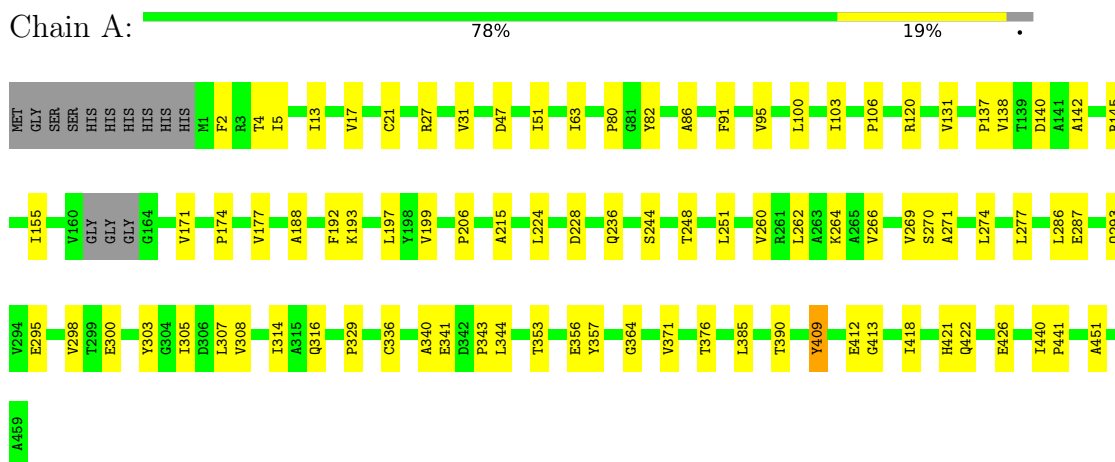
- Molecule 2 is a protein called Biotin carboxylase.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	E	23	Total	C	N	O	0	0	0
			182	117	36	29			
2	F	22	Total	C	N	O	0	0	0
			175	112	35	28			
2	G	19	Total	C	N	O	0	0	0
			154	99	32	23			
2	H	18	Total	C	N	O	0	0	0
			147	94	31	22			

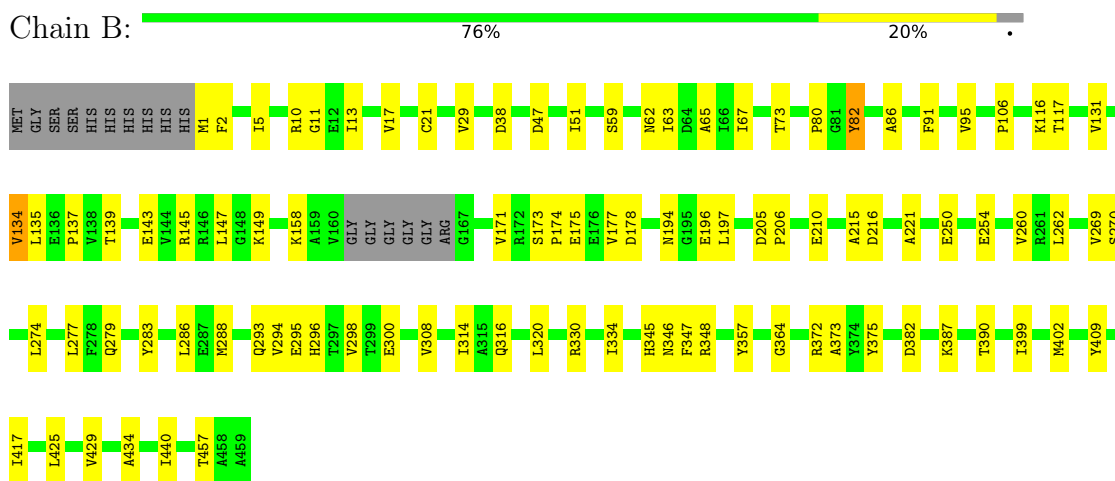
### 3 Residue-property plots [i](#)

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.

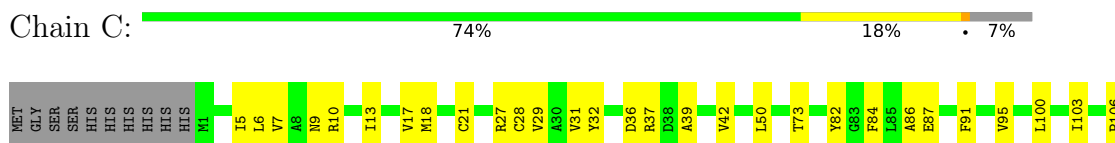
- Molecule 1: Biotin carboxylase

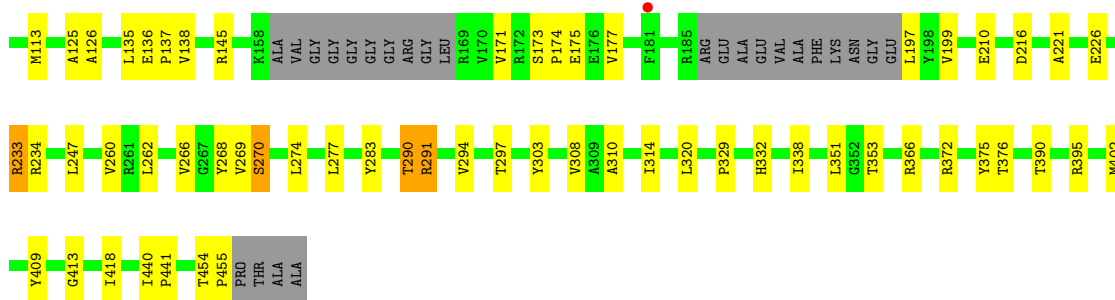


- Molecule 1: Biotin carboxylase

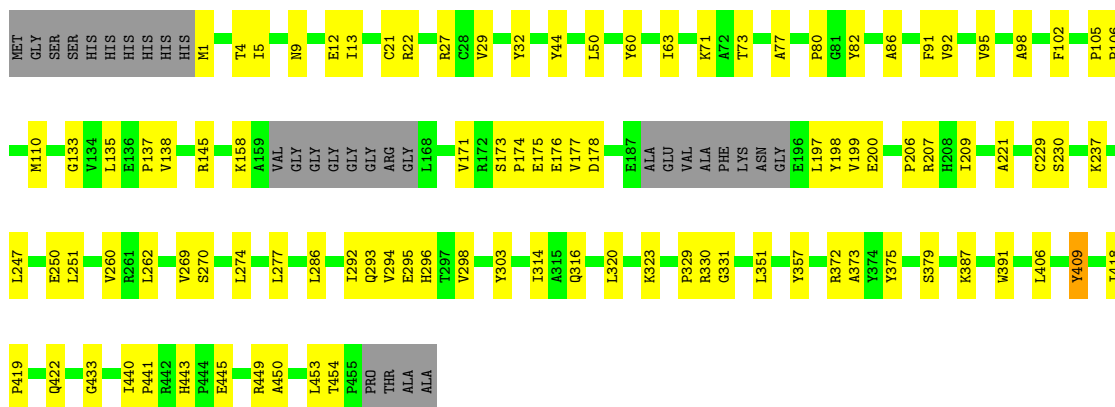


- Molecule 1: Biotin carboxylase

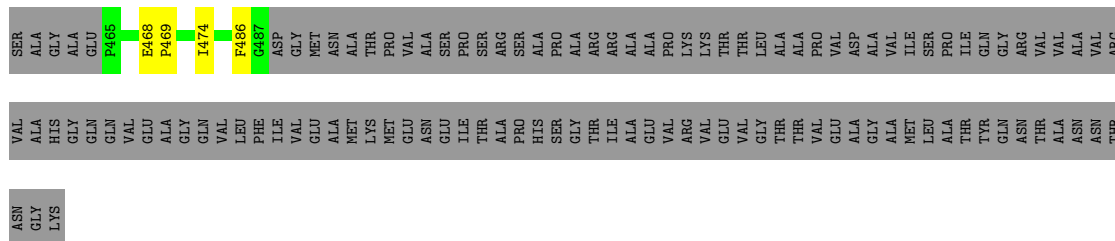




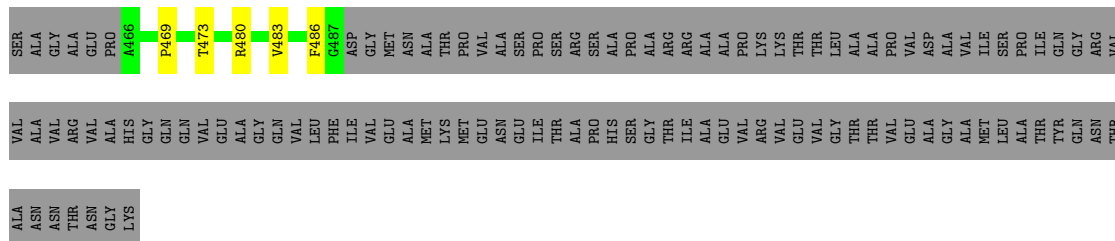
• Molecule 1: Biotin carboxylase



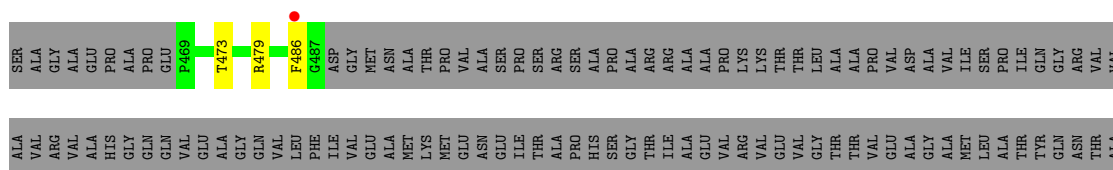
• Molecule 2: Biotin carboxylase



• Molecule 2: Biotin carboxylase

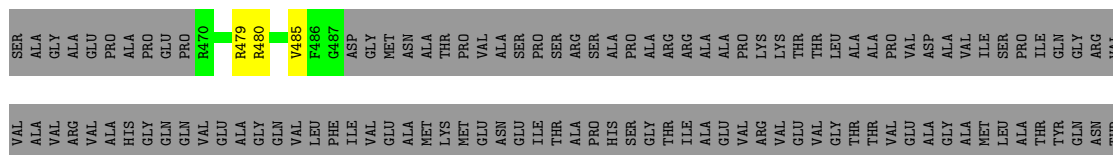


- Molecule 2: Biotin carboxylase



ASN	ASN	THR	ASN	GLY	LYS
-----	-----	-----	-----	-----	-----

- Molecule 2: Biotin carboxylase



ALA	ASN	ASN	THR	ASN	GLY	LYS
-----	-----	-----	-----	-----	-----	-----



## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	128.53Å 126.22Å 132.19Å 90.00° 106.44° 90.00°	Depositor
Resolution (Å)	28.80 – 3.20 28.80 – 3.20	Depositor EDS
% Data completeness (in resolution range)	99.4 (28.80-3.20) 99.3 (28.80-3.20)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.26 (at 3.17Å)	Xtrriage
Refinement program	PHENIX 1.19.2-4158	Depositor
R, $R_{free}$	0.212 , 0.244 0.215 , 0.242	Depositor DCC
$R_{free}$ test set	3184 reflections (4.79%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	74.8	Xtrriage
Anisotropy	0.089	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 35.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtrriage
Estimated twinning fraction	0.074 for l,-k,h	Xtrriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	14386	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	73.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.87% 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	0.34	0/3585	0.58	0/4870
1	B	0.40	0/3560	0.59	0/4838
1	C	0.33	0/3415	0.57	0/4640
1	D	0.33	0/3465	0.56	0/4706
2	E	0.30	0/187	0.62	0/251
2	F	0.41	0/179	0.63	0/240
2	G	0.30	0/157	0.68	0/208
2	H	0.37	0/149	0.55	0/197
All	All	0.35	0/14697	0.58	0/19950

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	3509	0	3483	66	0
1	B	3484	0	3460	55	0
1	C	3343	0	3310	62	0
1	D	3392	0	3362	69	0
2	E	182	0	182	3	0
2	F	175	0	174	2	0
2	G	154	0	157	2	0
2	H	147	0	149	1	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	14386	0	14277	256	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 9.

The worst 5 of 256 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:277:LEU:HG	1:D:286:LEU:HD11	1.19	1.12
1:D:106:PRO:HD3	1:D:269:VAL:HG22	1.32	1.05
1:D:4:THR:HG22	1:D:27:ARG:HB3	1.40	1.03
1:A:277:LEU:HG	1:A:286:LEU:HD11	1.43	1.00
1:A:270:SER:HA	1:A:314:ILE:HD11	1.42	0.98

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	452/469 (96%)	433 (96%)	19 (4%)	0	100	100
1	B	449/469 (96%)	428 (95%)	20 (4%)	1 (0%)	47	79
1	C	428/469 (91%)	412 (96%)	15 (4%)	1 (0%)	47	79
1	D	433/469 (92%)	419 (97%)	13 (3%)	1 (0%)	47	79
2	E	21/137 (15%)	21 (100%)	0	0	100	100
2	F	20/137 (15%)	20 (100%)	0	0	100	100
2	G	17/137 (12%)	16 (94%)	1 (6%)	0	100	100
2	H	16/137 (12%)	16 (100%)	0	0	100	100
All	All	1836/2424 (76%)	1765 (96%)	68 (4%)	3 (0%)	47	79

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	346	ASN
1	C	270	SER
1	D	60	TYR

### 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	349/360 (97%)	344 (99%)	5 (1%)	67 86
1	B	347/360 (96%)	337 (97%)	10 (3%)	42 74
1	C	333/360 (92%)	327 (98%)	6 (2%)	59 82
1	D	339/360 (94%)	335 (99%)	4 (1%)	71 88
2	E	18/102 (18%)	18 (100%)	0	100 100
2	F	17/102 (17%)	15 (88%)	2 (12%)	5 23
2	G	15/102 (15%)	14 (93%)	1 (7%)	16 50
2	H	14/102 (14%)	12 (86%)	2 (14%)	3 15
All	All	1432/1848 (78%)	1402 (98%)	30 (2%)	53 79

5 of 30 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	457	THR
2	G	479	ARG
1	C	290	THR
2	H	480	ARG
1	D	409	TYR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	289	ASN
1	B	293	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	C	232	GLN
1	C	289	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 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	456/469 (97%)	0.01	0 100 100	50, 67, 93, 128	0
1	B	453/469 (96%)	0.00	0 100 100	48, 70, 94, 122	0
1	C	434/469 (92%)	-0.01	1 (0%) 95 94	51, 69, 98, 132	0
1	D	439/469 (93%)	0.03	0 100 100	53, 73, 103, 142	0
2	E	23/137 (16%)	0.01	0 100 100	85, 101, 116, 127	0
2	F	22/137 (16%)	0.10	0 100 100	30, 100, 122, 127	0
2	G	19/137 (13%)	0.54	1 (5%) 26 14	75, 89, 124, 127	0
2	H	18/137 (13%)	0.36	0 100 100	84, 91, 109, 112	0
All	All	1864/2424 (76%)	0.02	2 (0%) 95 95	30, 71, 102, 142	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	181	PHE	3.0
2	G	486	PHE	2.5

### 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.