



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 14, 2020 – 04:44 am BST

PDB ID : 1Y2M  
Title : Crystal structure of phenylalanine ammonia-lyase from yeast *Rhododporidium toruloides*  
Authors : Wang, L.; Gamez, A.; Sarkissian, C.N.; Straub, M.; Patch, M.G.; Han, G.W.; Sriver, C.R.; Stevens, R.C.  
Deposited on : 2004-11-22  
Resolution : 1.60 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  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.11  
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) : 2.11

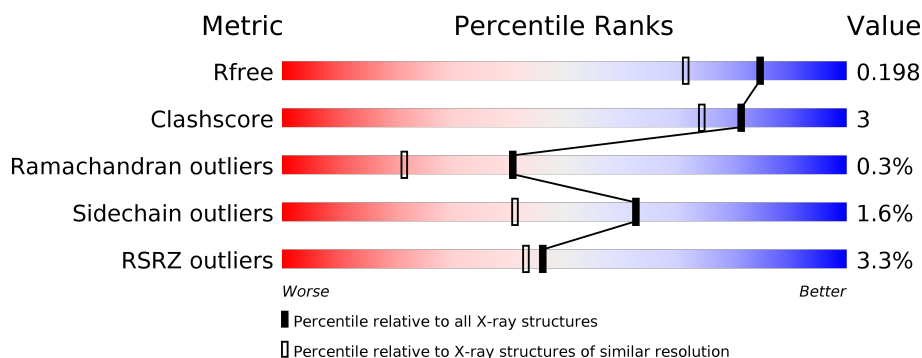
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.60 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	3398 (1.60-1.60)
Clashscore	141614	3665 (1.60-1.60)
Ramachandran outliers	138981	3564 (1.60-1.60)
Sidechain outliers	138945	3563 (1.60-1.60)
RSRZ outliers	127900	3321 (1.60-1.60)

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 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	716	<div> <div>3%</div> <div>87%</div> <div>5% • 7%</div> </div>
1	B	716	<div> <div>4%</div> <div>87%</div> <div>5% • 7%</div> </div>
1	C	716	<div> <div>2%</div> <div>87%</div> <div>5% 7%</div> </div>
1	D	716	<div> <div>3%</div> <div>87%</div> <div>5% 8%</div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 22123 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 Phenylalanine ammonia-lyase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	665	Total	C	N	O	S	Se	0	1	0
			5053	3177	896	963	4	13			
1	B	669	Total	C	N	O	S	Se	0	2	0
			5083	3195	899	971	4	14			
1	C	664	Total	C	N	O	S	Se	0	3	0
			5050	3175	897	960	4	14			
1	D	661	Total	C	N	O	S	Se	0	2	0
			5022	3158	888	958	4	14			

There are 60 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	MET	MODIFIED RESIDUE	UNP P11544
A	51	MSE	MET	MODIFIED RESIDUE	UNP P11544
A	107	MSE	MET	MODIFIED RESIDUE	UNP P11544
A	169	MSE	MET	MODIFIED RESIDUE	UNP P11544
A	250	MSE	MET	MODIFIED RESIDUE	UNP P11544
A	278	MSE	MET	MODIFIED RESIDUE	UNP P11544
A	288	MSE	MET	MODIFIED RESIDUE	UNP P11544
A	299	MSE	MET	MODIFIED RESIDUE	UNP P11544
A	304	MSE	MET	MODIFIED RESIDUE	UNP P11544
A	422	MSE	MET	MODIFIED RESIDUE	UNP P11544
A	443	MSE	MET	MODIFIED RESIDUE	UNP P11544
A	448	MSE	MET	MODIFIED RESIDUE	UNP P11544
A	497	MSE	MET	MODIFIED RESIDUE	UNP P11544
A	564	MSE	MET	MODIFIED RESIDUE	UNP P11544
A	714	MSE	MET	MODIFIED RESIDUE	UNP P11544
B	1	MSE	MET	MODIFIED RESIDUE	UNP P11544
B	51	MSE	MET	MODIFIED RESIDUE	UNP P11544
B	107	MSE	MET	MODIFIED RESIDUE	UNP P11544
B	169	MSE	MET	MODIFIED RESIDUE	UNP P11544
B	250	MSE	MET	MODIFIED RESIDUE	UNP P11544
B	278	MSE	MET	MODIFIED RESIDUE	UNP P11544

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	288	MSE	MET	MODIFIED RESIDUE	UNP P11544
B	299	MSE	MET	MODIFIED RESIDUE	UNP P11544
B	304	MSE	MET	MODIFIED RESIDUE	UNP P11544
B	422	MSE	MET	MODIFIED RESIDUE	UNP P11544
B	443	MSE	MET	MODIFIED RESIDUE	UNP P11544
B	448	MSE	MET	MODIFIED RESIDUE	UNP P11544
B	497	MSE	MET	MODIFIED RESIDUE	UNP P11544
B	564	MSE	MET	MODIFIED RESIDUE	UNP P11544
B	714	MSE	MET	MODIFIED RESIDUE	UNP P11544
C	1	MSE	MET	MODIFIED RESIDUE	UNP P11544
C	51	MSE	MET	MODIFIED RESIDUE	UNP P11544
C	107	MSE	MET	MODIFIED RESIDUE	UNP P11544
C	169	MSE	MET	MODIFIED RESIDUE	UNP P11544
C	250	MSE	MET	MODIFIED RESIDUE	UNP P11544
C	278	MSE	MET	MODIFIED RESIDUE	UNP P11544
C	288	MSE	MET	MODIFIED RESIDUE	UNP P11544
C	299	MSE	MET	MODIFIED RESIDUE	UNP P11544
C	304	MSE	MET	MODIFIED RESIDUE	UNP P11544
C	422	MSE	MET	MODIFIED RESIDUE	UNP P11544
C	443	MSE	MET	MODIFIED RESIDUE	UNP P11544
C	448	MSE	MET	MODIFIED RESIDUE	UNP P11544
C	497	MSE	MET	MODIFIED RESIDUE	UNP P11544
C	564	MSE	MET	MODIFIED RESIDUE	UNP P11544
C	714	MSE	MET	MODIFIED RESIDUE	UNP P11544
D	1	MSE	MET	MODIFIED RESIDUE	UNP P11544
D	51	MSE	MET	MODIFIED RESIDUE	UNP P11544
D	107	MSE	MET	MODIFIED RESIDUE	UNP P11544
D	169	MSE	MET	MODIFIED RESIDUE	UNP P11544
D	250	MSE	MET	MODIFIED RESIDUE	UNP P11544
D	278	MSE	MET	MODIFIED RESIDUE	UNP P11544
D	288	MSE	MET	MODIFIED RESIDUE	UNP P11544
D	299	MSE	MET	MODIFIED RESIDUE	UNP P11544
D	304	MSE	MET	MODIFIED RESIDUE	UNP P11544
D	422	MSE	MET	MODIFIED RESIDUE	UNP P11544
D	443	MSE	MET	MODIFIED RESIDUE	UNP P11544
D	448	MSE	MET	MODIFIED RESIDUE	UNP P11544
D	497	MSE	MET	MODIFIED RESIDUE	UNP P11544
D	564	MSE	MET	MODIFIED RESIDUE	UNP P11544
D	714	MSE	MET	MODIFIED RESIDUE	UNP P11544

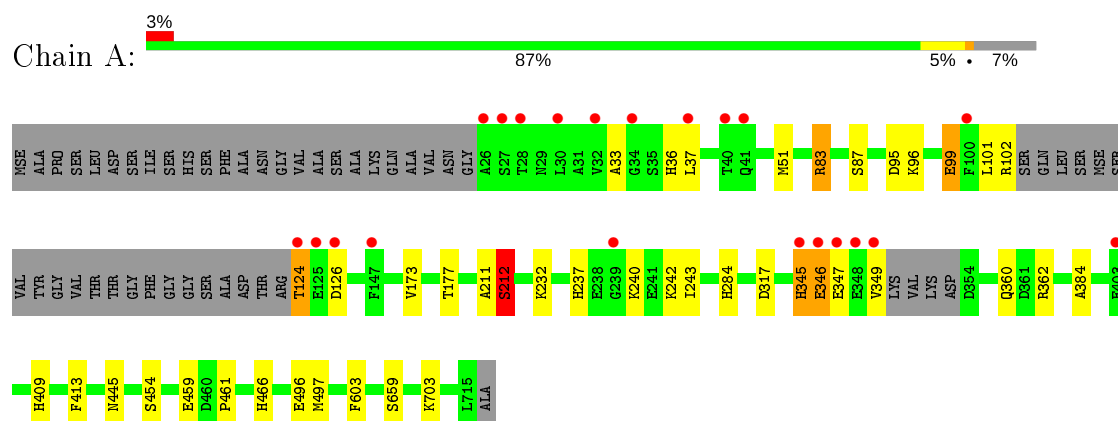
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	424	Total 424	O 424	0	0
2	B	537	Total 537	O 537	0	0
2	C	468	Total 468	O 468	0	1
2	D	486	Total 486	O 486	0	0

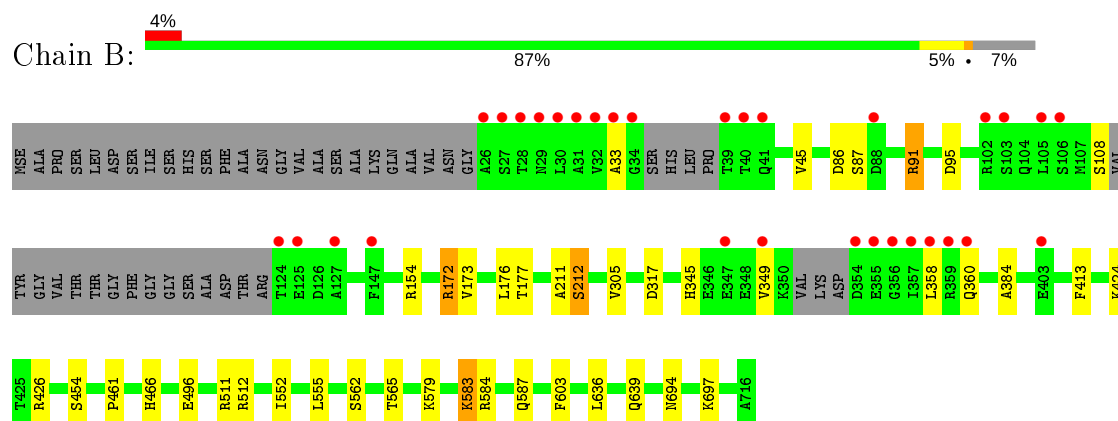
### 3 Residue-property plots [i](#)

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.

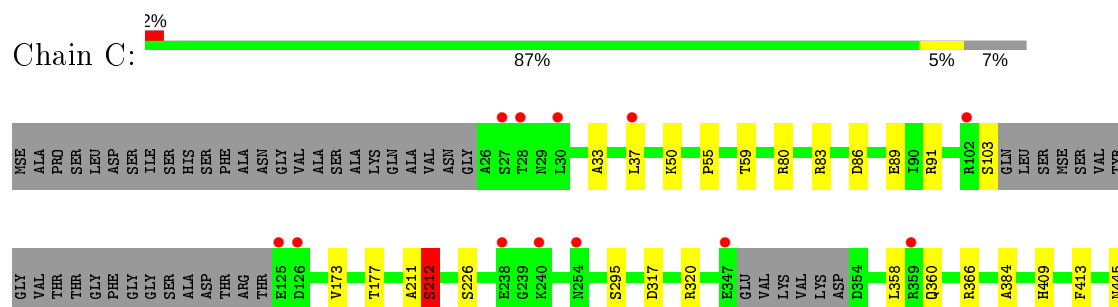
#### • Molecule 1: Phenylalanine ammonia-lyase



#### • Molecule 1: Phenylalanine ammonia-lyase

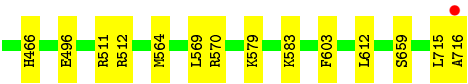
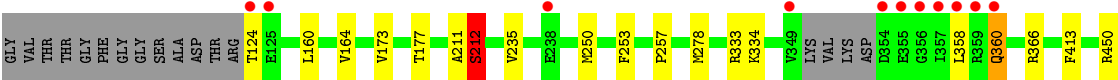
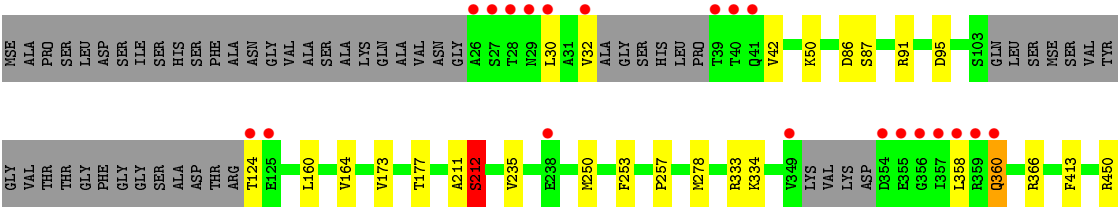
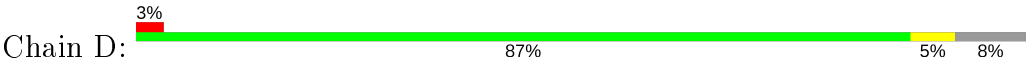


#### • Molecule 1: Phenylalanine ammonia-lyase





● Molecule 1: Phenylalanine ammonia-lyase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	104.76Å 151.61Å 179.92Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	41.88 – 1.60 41.33 – 1.60	Depositor EDS
% Data completeness (in resolution range)	99.2 (41.88-1.60) 99.2 (41.33-1.60)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.66 (at 1.60Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, $R_{free}$	0.158 , 0.187 0.171 , 0.198	Depositor DCC
$R_{free}$ test set	18674 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	14.1	Xtriage
Anisotropy	0.029	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 46.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	22123	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	14.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.28% 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	0.72	0/5126	0.77	4/6940 (0.1%)
1	B	0.76	0/5152	0.85	9/6967 (0.1%)
1	C	0.73	1/5132 (0.0%)	0.79	4/6943 (0.1%)
1	D	0.75	0/5097	0.81	8/6897 (0.1%)
All	All	0.74	1/20507 (0.0%)	0.81	25/27747 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	226	SER	CB-OG	-5.31	1.35	1.42

The worst 5 of 25 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	172	ARG	NE-CZ-NH1	12.85	126.72	120.30
1	B	172	ARG	NE-CZ-NH2	-12.28	114.16	120.30
1	C	320	ARG	NE-CZ-NH1	8.16	124.38	120.30
1	B	511	ARG	NE-CZ-NH2	-7.25	116.67	120.30
1	B	154	ARG	NE-CZ-NH1	7.12	123.86	120.30

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	5053	0	5106	26	0
1	B	5083	0	5142	25	0
1	C	5050	0	5108	20	0
1	D	5022	0	5076	43	0
2	A	424	0	0	8	0
2	B	537	0	0	6	0
2	C	468	0	0	5	0
2	D	486	0	0	2	0
All	All	22123	0	20432	107	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 107 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:564:MSE:HE3	1:D:612:LEU:HD21	1.47	0.96
1:C:295:SER:OG	1:C:714:MSE:HE1	1.69	0.93
1:A:36:HIS:HB2	2:A:984:HOH:O	1.66	0.92
1:D:42:VAL:HG13	1:D:42:VAL:O	1.71	0.90
1:D:164:VAL:HG21	1:D:278:MSE:HE2	1.59	0.83

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	660/716 (92%)	642 (97%)	15 (2%)	3 (0%)	29 11
1	B	663/716 (93%)	650 (98%)	11 (2%)	2 (0%)	41 21
1	C	661/716 (92%)	646 (98%)	13 (2%)	2 (0%)	41 21
1	D	655/716 (92%)	639 (98%)	14 (2%)	2 (0%)	41 21

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	2639/2864 (92%)	2577 (98%)	53 (2%)	9 (0%)	41 21

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	212	SER
1	B	212	SER
1	C	212	SER
1	C	466	HIS
1	D	212	SER

### 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	550/574 (96%)	537 (98%)	13 (2%)	49 24
1	B	554/574 (96%)	545 (98%)	9 (2%)	62 41
1	C	550/574 (96%)	543 (99%)	7 (1%)	69 50
1	D	547/574 (95%)	540 (99%)	7 (1%)	69 50
All	All	2201/2296 (96%)	2165 (98%)	36 (2%)	62 41

5 of 36 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	358	LEU
1	B	583	LYS
1	D	496	GLU
1	B	496	GLU
1	B	603	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	307	HIS
1	B	694	ASN
1	C	687	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	652/716 (91%)	-0.21	21 (3%) 47 44	6, 11, 27, 91	0
1	B	655/716 (91%)	-0.12	31 (4%) 31 28	6, 11, 30, 78	0
1	C	651/716 (90%)	-0.27	13 (1%) 65 64	7, 11, 27, 54	0
1	D	648/716 (90%)	-0.22	21 (3%) 47 44	5, 11, 28, 77	0
All	All	2606/2864 (90%)	-0.20	86 (3%) 46 43	5, 11, 28, 91	0

The worst 5 of 86 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	349	VAL	22.8
1	D	357	ILE	18.8
1	B	358	LEU	18.4
1	B	357	ILE	17.2
1	D	358	LEU	11.5

### 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 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.