



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 31, 2016 – 11:29 PM GMT

PDB ID : 1X8M  
Title : X-ray structure of pectin degrading enzyme 5-keto 4-deoxyuronate isomerase from Escherichia coli  
Authors : Fedorov, A.A.; Fedorov, E.V.; Almo, S.C.; Burley, S.K.; New York SGX Research Center for Structural Genomics (NYSGXRC)  
Deposited on : 2004-08-18  
Resolution : 2.60 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  
<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 : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
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) : trunk26865

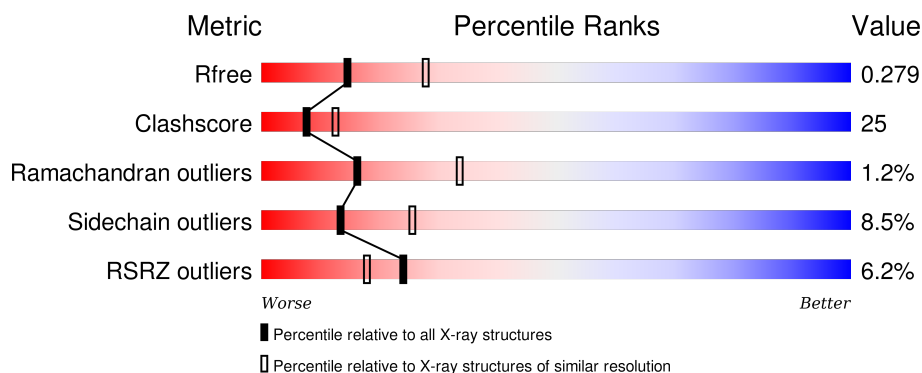
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.60 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))
$R_{free}$	91344	2328 (2.60-2.60)
Clashscore	102246	2679 (2.60-2.60)
Ramachandran outliers	100387	2635 (2.60-2.60)
Sidechain outliers	100360	2635 (2.60-2.60)
RSRZ outliers	91569	2334 (2.60-2.60)

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. 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	288	<div> <div>4%</div> <div>53% 31% 5% 10%</div> </div>
1	B	288	<div> <div>5%</div> <div>57% 28% 5% 10%</div> </div>
1	C	288	<div> <div>5%</div> <div>54% 31% 5% 10%</div> </div>
1	D	288	<div> <div>5%</div> <div>55% 29% 6% 10%</div> </div>
1	E	288	<div> <div>8%</div> <div>54% 30% 6% 10%</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	288	<div><div></div><div>7%</div><div>56%</div><div>28%</div><div>5%</div><div>10%</div></div>

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 12126 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 4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	259	Total	C	N	O	S	0	0	0
			2021	1284	341	383	13			
1	B	259	Total	C	N	O	S	0	0	0
			2021	1284	341	383	13			
1	C	259	Total	C	N	O	S	0	0	0
			2021	1284	341	383	13			
1	D	259	Total	C	N	O	S	0	0	0
			2021	1284	341	383	13			
1	E	259	Total	C	N	O	S	0	0	0
			2021	1284	341	383	13			
1	F	259	Total	C	N	O	S	0	0	0
			2021	1284	341	383	13			

There are 66 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1001	MET	-	CLONING ARTIFACT	UNP Q46938
A	1002	SER	-	CLONING ARTIFACT	UNP Q46938
A	1003	LEU	-	CLONING ARTIFACT	UNP Q46938
A	1281	GLU	-	EXPRESSION TAG	UNP Q46938
A	1282	GLY	-	EXPRESSION TAG	UNP Q46938
A	1283	HIS	-	EXPRESSION TAG	UNP Q46938
A	1284	HIS	-	EXPRESSION TAG	UNP Q46938
A	1285	HIS	-	EXPRESSION TAG	UNP Q46938
A	1286	HIS	-	EXPRESSION TAG	UNP Q46938
A	1287	HIS	-	EXPRESSION TAG	UNP Q46938
A	1288	HIS	-	EXPRESSION TAG	UNP Q46938
B	2001	MET	-	CLONING ARTIFACT	UNP Q46938
B	2002	SER	-	CLONING ARTIFACT	UNP Q46938
B	2003	LEU	-	CLONING ARTIFACT	UNP Q46938
B	2281	GLU	-	EXPRESSION TAG	UNP Q46938
B	2282	GLY	-	EXPRESSION TAG	UNP Q46938
B	2283	HIS	-	EXPRESSION TAG	UNP Q46938

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	2284	HIS	-	EXPRESSION TAG	UNP Q46938
B	2285	HIS	-	EXPRESSION TAG	UNP Q46938
B	2286	HIS	-	EXPRESSION TAG	UNP Q46938
B	2287	HIS	-	EXPRESSION TAG	UNP Q46938
B	2288	HIS	-	EXPRESSION TAG	UNP Q46938
C	3001	MET	-	CLONING ARTIFACT	UNP Q46938
C	3002	SER	-	CLONING ARTIFACT	UNP Q46938
C	3003	LEU	-	CLONING ARTIFACT	UNP Q46938
C	3281	GLU	-	EXPRESSION TAG	UNP Q46938
C	3282	GLY	-	EXPRESSION TAG	UNP Q46938
C	3283	HIS	-	EXPRESSION TAG	UNP Q46938
C	3284	HIS	-	EXPRESSION TAG	UNP Q46938
C	3285	HIS	-	EXPRESSION TAG	UNP Q46938
C	3286	HIS	-	EXPRESSION TAG	UNP Q46938
C	3287	HIS	-	EXPRESSION TAG	UNP Q46938
C	3288	HIS	-	EXPRESSION TAG	UNP Q46938
D	4001	MET	-	CLONING ARTIFACT	UNP Q46938
D	4002	SER	-	CLONING ARTIFACT	UNP Q46938
D	4003	LEU	-	CLONING ARTIFACT	UNP Q46938
D	4281	GLU	-	EXPRESSION TAG	UNP Q46938
D	4282	GLY	-	EXPRESSION TAG	UNP Q46938
D	4283	HIS	-	EXPRESSION TAG	UNP Q46938
D	4284	HIS	-	EXPRESSION TAG	UNP Q46938
D	4285	HIS	-	EXPRESSION TAG	UNP Q46938
D	4286	HIS	-	EXPRESSION TAG	UNP Q46938
D	4287	HIS	-	EXPRESSION TAG	UNP Q46938
D	4288	HIS	-	EXPRESSION TAG	UNP Q46938
E	5001	MET	-	CLONING ARTIFACT	UNP Q46938
E	5002	SER	-	CLONING ARTIFACT	UNP Q46938
E	5003	LEU	-	CLONING ARTIFACT	UNP Q46938
E	5281	GLU	-	EXPRESSION TAG	UNP Q46938
E	5282	GLY	-	EXPRESSION TAG	UNP Q46938
E	5283	HIS	-	EXPRESSION TAG	UNP Q46938
E	5284	HIS	-	EXPRESSION TAG	UNP Q46938
E	5285	HIS	-	EXPRESSION TAG	UNP Q46938
E	5286	HIS	-	EXPRESSION TAG	UNP Q46938
E	5287	HIS	-	EXPRESSION TAG	UNP Q46938
E	5288	HIS	-	EXPRESSION TAG	UNP Q46938
F	6001	MET	-	CLONING ARTIFACT	UNP Q46938
F	6002	SER	-	CLONING ARTIFACT	UNP Q46938
F	6003	LEU	-	CLONING ARTIFACT	UNP Q46938
F	6281	GLU	-	EXPRESSION TAG	UNP Q46938

*Continued on next page...*

*Continued from previous page...*


Chain	Residue	Modelled	Actual	Comment	Reference
F	6282	GLY	-	EXPRESSION TAG	UNP Q46938
F	6283	HIS	-	EXPRESSION TAG	UNP Q46938
F	6284	HIS	-	EXPRESSION TAG	UNP Q46938
F	6285	HIS	-	EXPRESSION TAG	UNP Q46938
F	6286	HIS	-	EXPRESSION TAG	UNP Q46938
F	6287	HIS	-	EXPRESSION TAG	UNP Q46938
F	6288	HIS	-	EXPRESSION TAG	UNP Q46938

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.

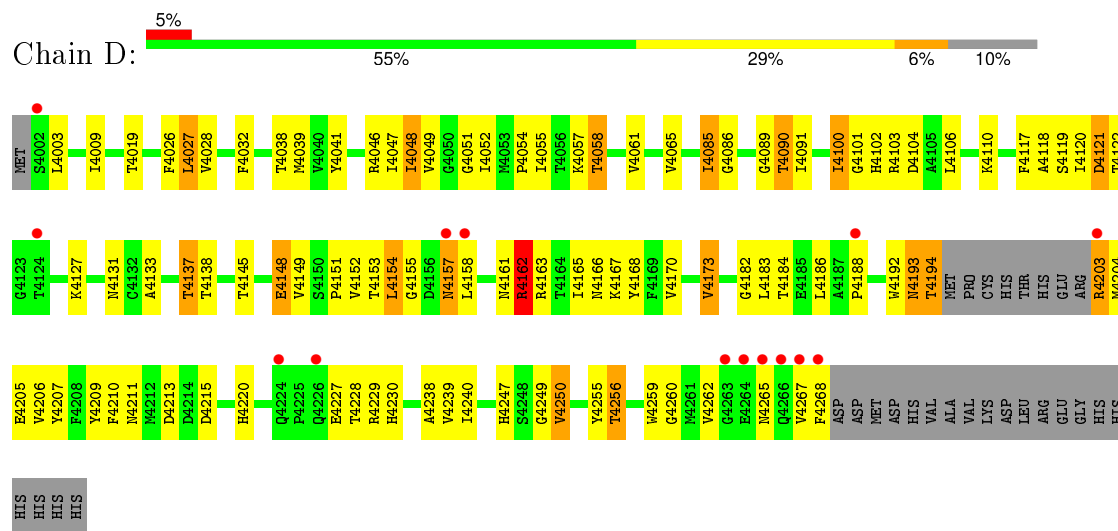
- Chain A:

Amino Acid	Type	Percentage
MET	Grey	4%
L1002	Green	
A1034	Green	
A1038	Green	
A1040	Green	
A1041	Green	
A1044	Green	
A1049	Green	
A1051	Green	
A1052	Green	
A1053	Green	
A1054	Green	
A1055	Green	
A1056	Green	
A1057	Green	
A1058	Green	
A1061	Green	
A1065	Green	
A1066	Green	
A1067	Green	
A1070	Green	
A1085	Green	
A1086	Green	
A1087	Green	
A1088	Green	
A1089	Green	
A1090	Green	
A1091	Green	
A1100	Green	
A1101	Green	
A1102	Green	
A1103	Green	

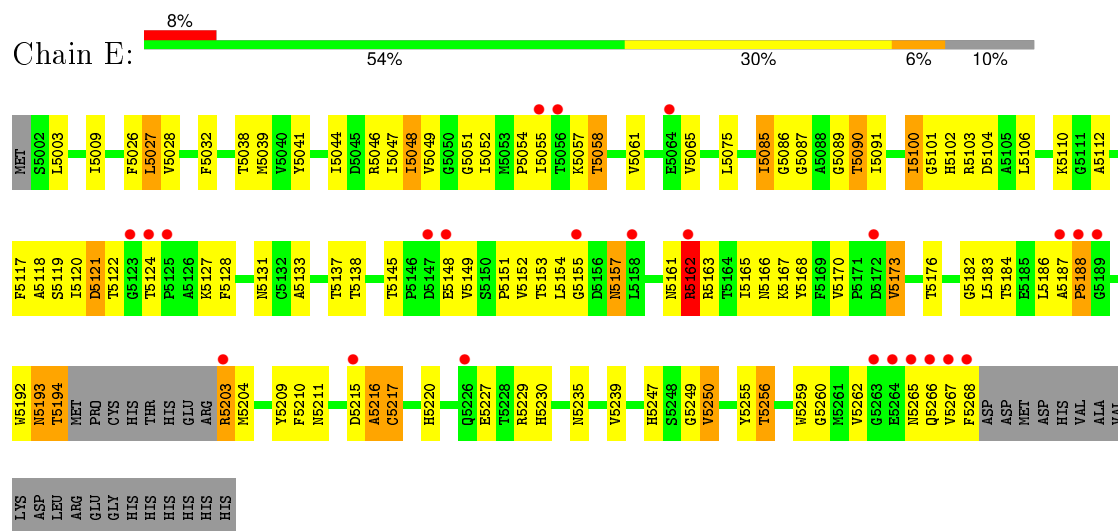
- [illegible]

- Chain C: 
- | MET   |
|-------|
| S3002 |
| L3003 |
| Q3007 |
| S3008 |
| L3009 |
| H3010 |
| S3011 |
| F3026 |
| S3027 |
| L3028 |
| F3032 |
| M3039 |
| V3040 |
| Y3041 |
| R3046 |
| L3047 |
| L3048 |
| V3049 |
| S3052 |
| L3053 |
| P3054 |
| L3055 |
| S3056 |
| K3057 |
| T3058 |
| V3061 |
| V3065 |
| L3075 |
| E3076 |
| R3077 |
| L3085 |
| G3086 |
| G3089 |
| T3090 |
| L3091 |
| I3100 |
| G3101 |
| H3102 |
| R3103 |
| D3104 |
| A3105 |
| L3106 |
| K3110 |
| G3111 |
| A3112 |

- Molecule 1: 4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase



- Molecule 1: 4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase



- Molecule 1: 4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase







## 4 Data and refinement statistics

Property	Value	Source
Space group	P 31	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	101.01Å 101.01Å 179.41Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	25.00 – 2.60 29.25 – 2.60	Depositor EDS
% Data completeness (in resolution range)	99.1 (25.00-2.60) 99.1 (29.25-2.60)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.12 (at 2.61Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, $R_{free}$	0.257 , 0.280 0.254 , 0.279	Depositor DCC
$R_{free}$ test set	3158 reflections (5.07%)	DCC
Wilson B-factor (Å <sup>2</sup> )	41.4	Xtriage
Anisotropy	0.325	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 43.1	EDS
Estimated twinning fraction	0.000 for -h,-k,l 0.037 for h,-h-k,-l 0.000 for -k,-h,-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 62822 reflections	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	12126	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 44.77 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 1.4544e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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	0.42	0/2067	0.68	0/2806
1	B	0.42	0/2067	0.67	0/2806
1	C	0.45	0/2067	0.67	0/2806
1	D	0.45	0/2067	0.67	0/2806
1	E	0.45	0/2067	0.69	0/2806
1	F	0.42	0/2067	0.67	0/2806
All	All	0.43	0/12402	0.67	0/16836

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

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	2021	0	1972	116	0
1	B	2021	0	1972	93	0
1	C	2021	0	1972	112	0
1	D	2021	0	1972	112	0
1	E	2021	0	1972	102	0
1	F	2021	0	1972	104	0
All	All	12126	0	11832	607	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 25.

The worst 5 of 607 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:E:5163:ARG:HB3	1:E:5186:LEU:HD23	1.36	1.08
1:D:4163:ARG:HB3	1:D:4186:LEU:HD23	1.37	1.07
1:C:3163:ARG:HB3	1:C:3186:LEU:HD23	1.37	1.07
1:A:1163:ARG:HB3	1:A:1186:LEU:HD23	1.36	1.05
1:F:6163:ARG:HB3	1:F:6186:LEU:HD23	1.38	1.04

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	255/288 (88%)	236 (92%)	16 (6%)	3 (1%)	16	33
1	B	255/288 (88%)	236 (92%)	17 (7%)	2 (1%)	24	46
1	C	255/288 (88%)	236 (92%)	16 (6%)	3 (1%)	16	33
1	D	255/288 (88%)	235 (92%)	16 (6%)	4 (2%)	12	24
1	E	255/288 (88%)	234 (92%)	17 (7%)	4 (2%)	12	24
1	F	255/288 (88%)	236 (92%)	16 (6%)	3 (1%)	16	33
All	All	1530/1728 (88%)	1413 (92%)	98 (6%)	19 (1%)	16	33

5 of 19 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	1121	ASP
1	A	1162	ARG
1	B	2121	ASP
1	B	2162	ARG
1	C	3121	ASP

### 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	221/248 (89%)	203 (92%)	18 (8%)	15	28
1	B	221/248 (89%)	201 (91%)	20 (9%)	12	23
1	C	221/248 (89%)	203 (92%)	18 (8%)	15	28
1	D	221/248 (89%)	202 (91%)	19 (9%)	13	25
1	E	221/248 (89%)	201 (91%)	20 (9%)	12	23
1	F	221/248 (89%)	203 (92%)	18 (8%)	15	28
All	All	1326/1488 (89%)	1213 (92%)	113 (8%)	13	25

5 of 113 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	3203	ARG
1	D	4148	GLU
1	F	6162	ARG
1	C	3239	VAL
1	D	4048	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 69 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	3220	HIS
1	D	4131	ASN
1	F	6161	ASN
1	C	3265	ASN
1	D	4024	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 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	259/288 (89%)	0.09	12 (4%) 36 29	13, 36, 75, 125	0
1	B	259/288 (89%)	0.07	14 (5%) 29 22	7, 32, 76, 124	0
1	C	259/288 (89%)	0.15	14 (5%) 29 22	10, 38, 83, 130	0
1	D	259/288 (89%)	0.13	14 (5%) 29 22	9, 38, 82, 129	0
1	E	259/288 (89%)	0.58	24 (9%) 11 7	16, 52, 87, 132	0
1	F	259/288 (89%)	0.37	19 (7%) 18 12	19, 45, 85, 129	0
All	All	1554/1728 (89%)	0.23	97 (6%) 24 18	7, 41, 83, 132	0

The worst 5 of 97 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	3267	VAL	8.6
1	C	3264	GLU	7.6
1	C	3265	ASN	6.7
1	F	6268	PHE	6.2
1	E	5268	PHE	6.2

### 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.