wwPDB X-ray Structure Validation Summary Report (i)

Nov 30, 2023 – 06:05 PM EST

PDB ID	:	8ETN
Title	:	The X-ray Crystal Structure of Tri-Ketone Dioxygenase from Rice
Authors	:	Rydel, T.J.; Duda, D.; Zheng, M.; Duff, S.M.G.
Deposited on	:	2022-10-17
Resolution	:	3.16 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) 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 (1)) were used in the production of this report:

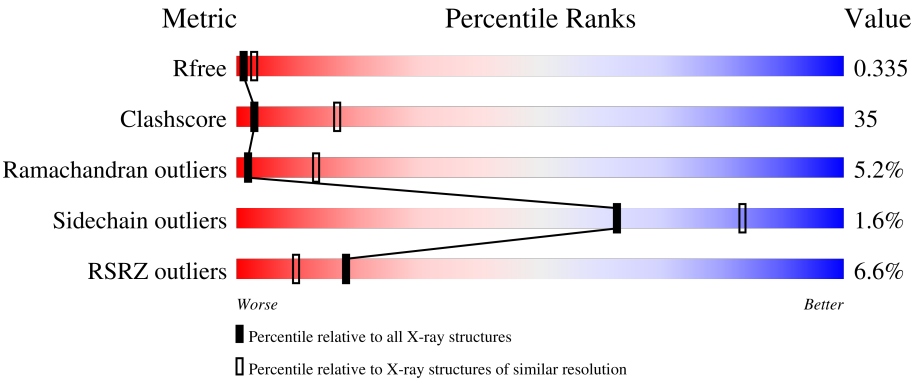
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 3.16 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	1665 (3.20-3.12)
Clashscore	141614	1804 (3.20-3.12)
Ramachandran outliers	138981	1770 (3.20-3.12)
Sidechain outliers	138945	1769 (3.20-3.12)
RSRZ outliers	127900	1616 (3.20-3.12)

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 >=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 <=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	360	7% 26% 36% 6% 32%
1	B	360	2% 27% 40% • • 29%
1	C	360	6% 27% 38% • 32%
1	D	360	3% 32% 35% • 29%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 8092 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 Tri-Ketone Dioxygenase.

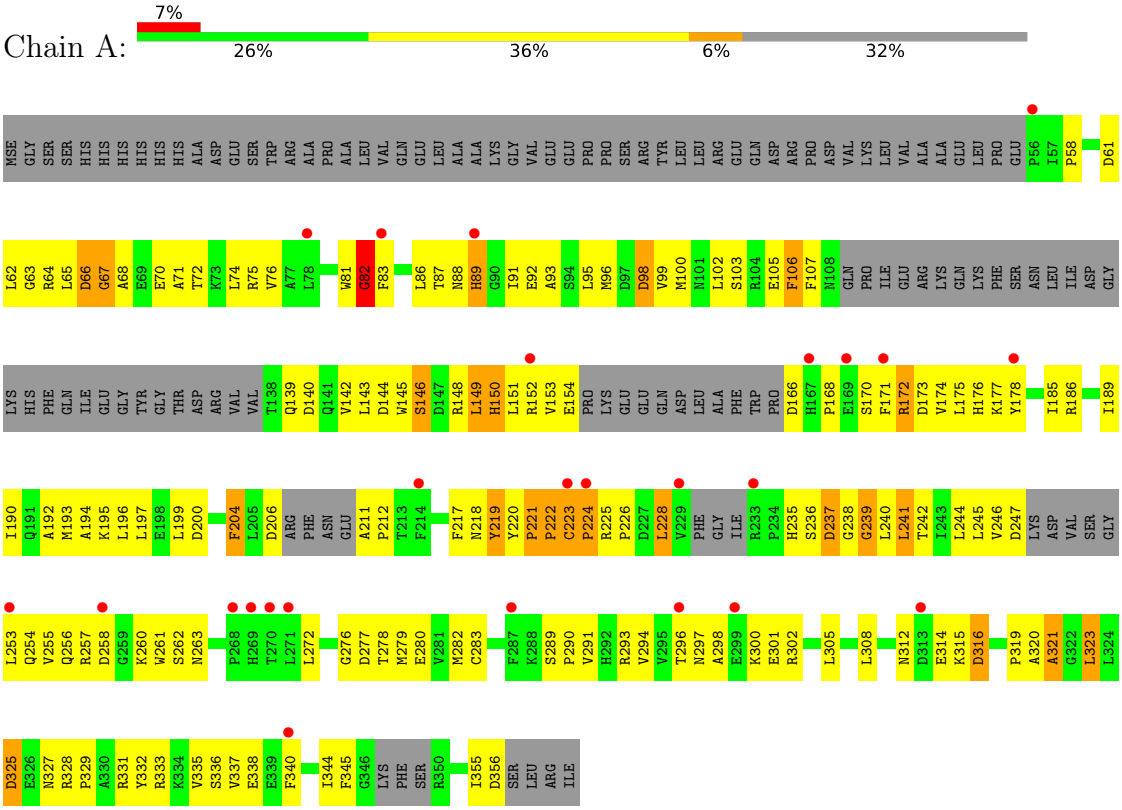
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	246	Total	C	N	O	S	Se	0	0	0
			1980	1252	349	370	3	6			
1	D	255	Total	C	N	O	S	Se	0	0	0
			2061	1313	360	379	3	6			
1	C	246	Total	C	N	O	S	Se	0	0	0
			1981	1256	349	367	3	6			
1	B	256	Total	C	N	O	S	Se	0	0	0
			2070	1318	362	381	3	6			

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.

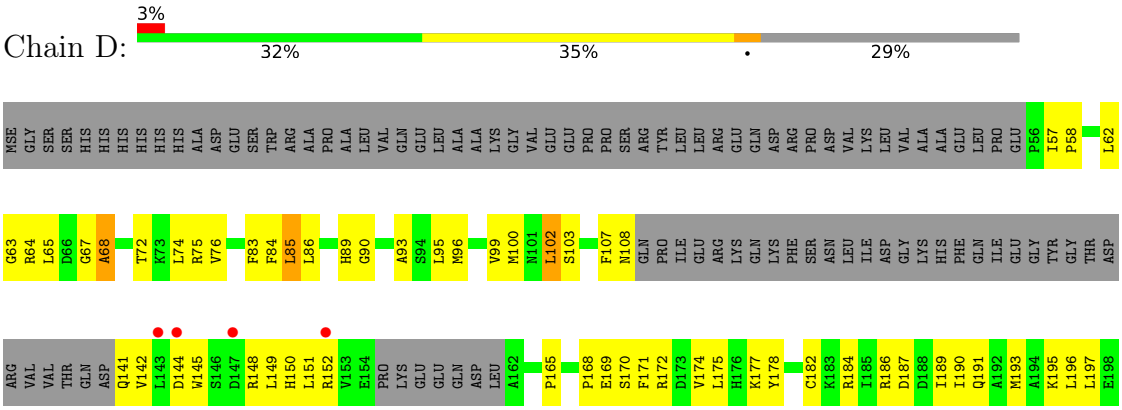
• Molecule 1: Tri-Ketone Dioxygenase

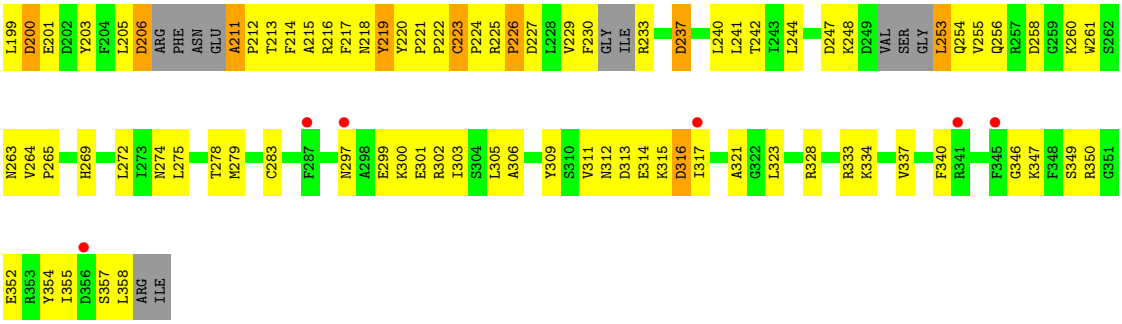
Chain A:



• Molecule 1: Tri-Ketone Dioxygenase

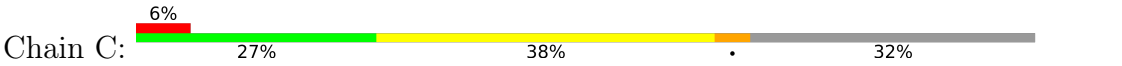
Chain D:

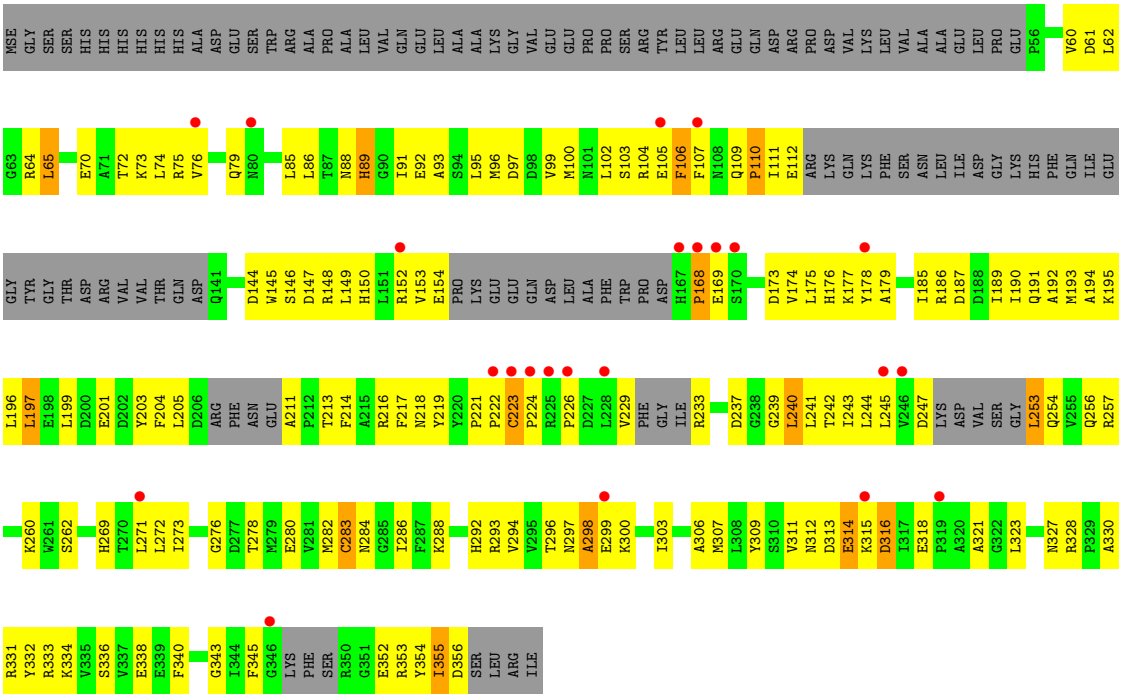




• Molecule 1: Tri-Ketone Dioxygenase

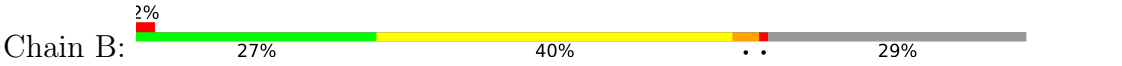
Chain C:

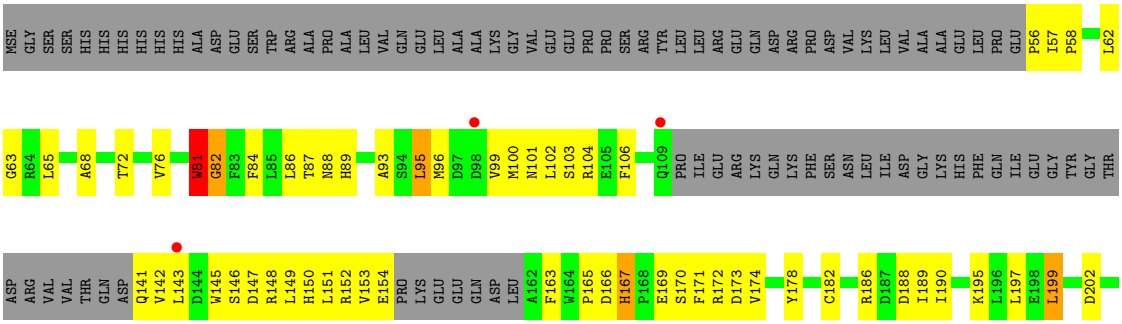


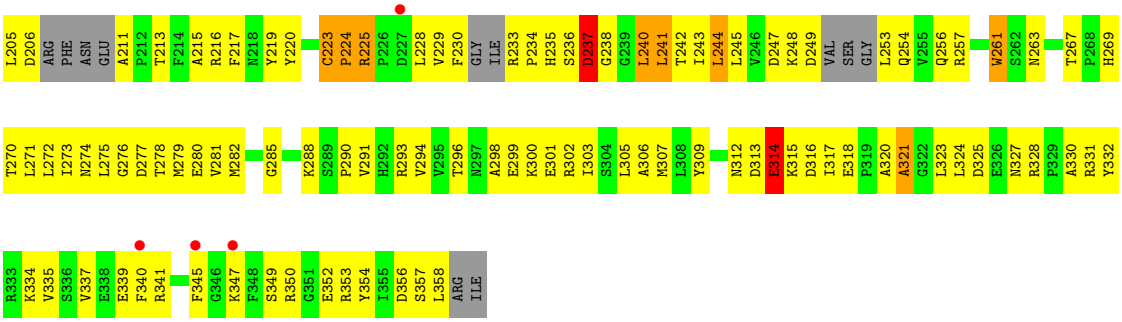


• Molecule 1: Tri-Ketone Dioxygenase

Chain B:






L205 D206 ARG PHE ASN GLU A211 P212 T213 F214 A215 R216 F217 N218 Y219 Y220 C223 P224 R225 P226 D227 L228 L229 V230 GLY ILE R233 P234 H235 S236 D237 G238 G239 L240 L241 T242 I243 L244 L245 V246 D247 K248 D249 VAL SER GLY L253 Q254 V255 Q256 R257 W261 S262 N263 T267 P268 H269
T270 L271 L272 I273 N274 L275 G276 D277 T278 M279 E280 V281 M282 G285 K288 S289 P290 V291 H292 R293 V294 V295 T296 N297 A298 E299 K300 E301 R302 I303 S304 L305 A306 M307 L308 Y309 N312 D313 E314 K315 D316 I317 E318 P319 A320 A321 G322 L323 L324 D325 E326 N327 R328 P329 A330 R331 Y332
R333 K334 V335 S336 V337 E338 E339 F340 R341 F345 G346 K347 F348 S349 R350 G351 E352 R353 Y354 I355 D356 S357 L358 ARG ILE

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41	Depositor
Cell constants a, b, c, α, β, γ	73.56Å 73.56Å 259.83Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	36.78 – 3.16 36.78 – 3.16	Depositor EDS
% Data completeness (in resolution range)	99.4 (36.78-3.16) 99.4 (36.78-3.16)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	0.15	Depositor
$< I/\sigma(I) >$ [1]	2.87 (at 3.18Å)	Xtriage
Refinement program	REFMAC 5, PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.274 , 0.318 0.289 , 0.335	Depositor DCC
R_{free} test set	1143 reflections (4.88%)	wwPDB-VP
Wilson B-factor (Å^2)	110.3	Xtriage
Anisotropy	0.376	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 67.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.468 for h,-k,-l	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	8092	wwPDB-VP
Average B, all atoms (Å^2)	110.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.54% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 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	0.74	1/2012 (0.0%)	1.07	4/2708 (0.1%)
1	B	0.79	2/2109 (0.1%)	1.06	8/2840 (0.3%)
1	C	0.72	1/2014 (0.0%)	1.08	9/2711 (0.3%)
1	D	0.79	2/2101 (0.1%)	1.01	7/2831 (0.2%)
All	All	0.76	6/8236 (0.1%)	1.06	28/11090 (0.3%)

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	1
1	B	0	3
1	C	0	1
1	D	0	1
All	All	0	6

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	261	TRP	CB-CG	-8.13	1.35	1.50
1	C	283	CYS	CB-SG	-7.18	1.70	1.82
1	D	283	CYS	CB-SG	-6.99	1.70	1.82
1	A	283	CYS	CB-SG	-6.91	1.70	1.82
1	B	81	TRP	CB-CG	-5.89	1.39	1.50

The worst 5 of 28 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	74	LEU	CA-CB-CG	7.80	133.23	115.30
1	C	245	LEU	CA-CB-CG	7.37	132.25	115.30
1	A	228	LEU	CA-CB-CG	7.33	132.15	115.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	85	LEU	CA-CB-CG	7.12	131.67	115.30
1	B	240	LEU	CA-CB-CG	7.04	131.49	115.30

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	211	ALA	Peptide
1	B	170	SER	Peptide
1	B	223	CYS	Peptide
1	C	355	ILE	Peptide
1	D	346	GLY	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	1980	0	1940	170	0
1	B	2070	0	2030	163	0
1	C	1981	0	1949	163	0
1	D	2061	0	2023	135	0
All	All	8092	0	7942	557	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 35.

The worst 5 of 557 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:B:256:GLN:CD	1:B:261:TRP:HE1	1.23	1.41
1:B:256:GLN:NE2	1:B:261:TRP:HE1	1.26	1.31
1:B:256:GLN:NE2	1:B:261:TRP:NE1	1.88	1.18
1:B:256:GLN:CD	1:B:261:TRP:NE1	2.07	1.07
1:B:253:LEU:HB3	1:B:302:ARG:HH22	1.23	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	232/360 (64%)	187 (81%)	23 (10%)	22 (10%)	0 3
1	B	244/360 (68%)	209 (86%)	27 (11%)	8 (3%)	4 22
1	C	232/360 (64%)	190 (82%)	31 (13%)	11 (5%)	2 15
1	D	245/360 (68%)	207 (84%)	29 (12%)	9 (4%)	3 19
All	All	953/1440 (66%)	793 (83%)	110 (12%)	50 (5%)	2 13

5 of 50 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	82	GLY
1	A	212	PRO
1	A	219	TYR
1	A	222	PRO
1	A	223	CYS

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	216/308 (70%)	210 (97%)	6 (3%)	43 73
1	B	225/308 (73%)	221 (98%)	4 (2%)	59 81
1	C	216/308 (70%)	213 (99%)	3 (1%)	67 85
1	D	224/308 (73%)	223 (100%)	1 (0%)	91 96
All	All	881/1232 (72%)	867 (98%)	14 (2%)	62 83

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	73	LYS
1	C	89	HIS
1	B	237	ASP
1	B	163	PHE
1	B	235	HIS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	150	HIS
1	C	284	ASN
1	B	256	GLN
1	B	108	ASN
1	C	79	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

6.1 Protein, DNA and RNA chains

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^{th} 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^2)	Q<0.9
1	A	240/360 (66%)	0.56	25 (10%) 6 3	81, 112, 152, 168	0
1	B	250/360 (69%)	0.22	7 (2%) 53 36	81, 102, 137, 156	0
1	C	240/360 (66%)	0.49	23 (9%) 8 4	81, 112, 153, 180	0
1	D	249/360 (69%)	0.30	10 (4%) 38 23	81, 101, 139, 159	0
All	All	979/1440 (67%)	0.39	65 (6%) 18 10	81, 107, 147, 180	0

The worst 5 of 65 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	224	PRO	11.8
1	C	225	ARG	8.4
1	A	171	PHE	6.7
1	C	223	CYS	6.4
1	A	287	PHE	6.3

6.2 Non-standard residues in protein, DNA, RNA chains

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

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.