W O R L D W I D E
wwPDB
P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Aug 20, 2023 – 12:41 AM EDT

PDB ID	:	2GPW
Title	:	Crystal Structure of the Biotin Carboxylase Subunit, F363A Mutant, of Acetyl-CoA Carboxylase from Escherichia coli.
Authors	:	Shen, Y.; Chou, C.Y.; Chang, G.G.; Tong, L.
Deposited on	:	2006-04-18
Resolution	:	2.20 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

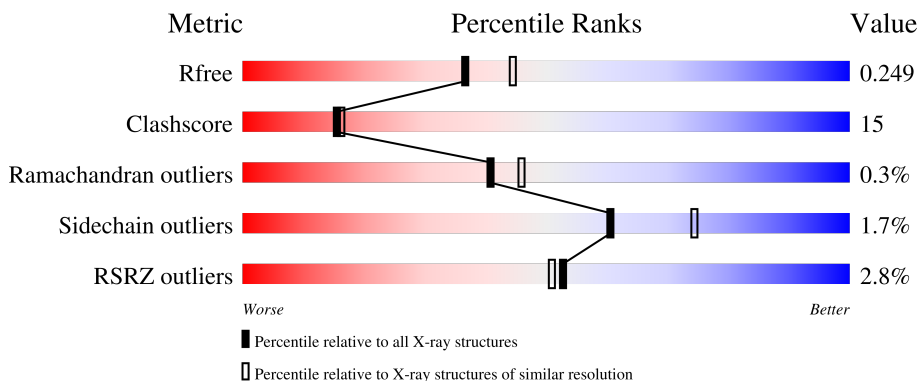|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 2.20 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	Percentile Ranks	Value
Rfree		0.249
Clashscore		15
Ramachandran outliers		0.3%
Sidechain outliers		1.7%
RSRZ outliers		2.8%

Worse   *Better*

▌ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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 >=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	469	3% 73% 21% • 5%
1	B	469	3% 73% 20% • 6%
1	C	469	2% 74% 20% • 5%
1	D	469	3% 68% 26% • 5%

2 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 14802 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
1	A	445	Total	C	N	O	S	0	0	0
			3434	2162	615	635	22			
1	B	442	Total	C	N	O	S	0	0	0
			3410	2146	611	631	22			
1	C	446	Total	C	N	O	S	0	0	0
			3440	2165	616	637	22			
1	D	447	Total	C	N	O	S	0	0	0
			3447	2170	617	638	22			

There are 84 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-19	MET	-	cloning artifact	UNP P24182
A	-18	GLY	-	cloning artifact	UNP P24182
A	-17	SER	-	cloning artifact	UNP P24182
A	-16	SER	-	cloning artifact	UNP P24182
A	-15	HIS	-	expression tag	UNP P24182
A	-14	HIS	-	expression tag	UNP P24182
A	-13	HIS	-	expression tag	UNP P24182
A	-12	HIS	-	expression tag	UNP P24182
A	-11	HIS	-	expression tag	UNP P24182
A	-10	HIS	-	expression tag	UNP P24182
A	-9	SER	-	cloning artifact	UNP P24182
A	-8	SER	-	cloning artifact	UNP P24182
A	-7	GLY	-	cloning artifact	UNP P24182
A	-6	LEU	-	cloning artifact	UNP P24182
A	-5	VAL	-	cloning artifact	UNP P24182
A	-4	PRO	-	cloning artifact	UNP P24182
A	-3	ARG	-	cloning artifact	UNP P24182
A	-2	GLY	-	cloning artifact	UNP P24182
A	-1	SER	-	cloning artifact	UNP P24182
A	0	HIS	-	cloning artifact	UNP P24182
A	363	ALA	PHE	engineered mutation	UNP P24182

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-19	MET	-	cloning artifact	UNP P24182
B	-18	GLY	-	cloning artifact	UNP P24182
B	-17	SER	-	cloning artifact	UNP P24182
B	-16	SER	-	cloning artifact	UNP P24182
B	-15	HIS	-	expression tag	UNP P24182
B	-14	HIS	-	expression tag	UNP P24182
B	-13	HIS	-	expression tag	UNP P24182
B	-12	HIS	-	expression tag	UNP P24182
B	-11	HIS	-	expression tag	UNP P24182
B	-10	HIS	-	expression tag	UNP P24182
B	-9	SER	-	cloning artifact	UNP P24182
B	-8	SER	-	cloning artifact	UNP P24182
B	-7	GLY	-	cloning artifact	UNP P24182
B	-6	LEU	-	cloning artifact	UNP P24182
B	-5	VAL	-	cloning artifact	UNP P24182
B	-4	PRO	-	cloning artifact	UNP P24182
B	-3	ARG	-	cloning artifact	UNP P24182
B	-2	GLY	-	cloning artifact	UNP P24182
B	-1	SER	-	cloning artifact	UNP P24182
B	0	HIS	-	cloning artifact	UNP P24182
B	363	ALA	PHE	engineered mutation	UNP P24182
C	-19	MET	-	cloning artifact	UNP P24182
C	-18	GLY	-	cloning artifact	UNP P24182
C	-17	SER	-	cloning artifact	UNP P24182
C	-16	SER	-	cloning artifact	UNP P24182
C	-15	HIS	-	expression tag	UNP P24182
C	-14	HIS	-	expression tag	UNP P24182
C	-13	HIS	-	expression tag	UNP P24182
C	-12	HIS	-	expression tag	UNP P24182
C	-11	HIS	-	expression tag	UNP P24182
C	-10	HIS	-	expression tag	UNP P24182
C	-9	SER	-	cloning artifact	UNP P24182
C	-8	SER	-	cloning artifact	UNP P24182
C	-7	GLY	-	cloning artifact	UNP P24182
C	-6	LEU	-	cloning artifact	UNP P24182
C	-5	VAL	-	cloning artifact	UNP P24182
C	-4	PRO	-	cloning artifact	UNP P24182
C	-3	ARG	-	cloning artifact	UNP P24182
C	-2	GLY	-	cloning artifact	UNP P24182
C	-1	SER	-	cloning artifact	UNP P24182
C	0	HIS	-	cloning artifact	UNP P24182
C	363	ALA	PHE	engineered mutation	UNP P24182

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	-19	MET	-	cloning artifact	UNP P24182
D	-18	GLY	-	cloning artifact	UNP P24182
D	-17	SER	-	cloning artifact	UNP P24182
D	-16	SER	-	cloning artifact	UNP P24182
D	-15	HIS	-	expression tag	UNP P24182
D	-14	HIS	-	expression tag	UNP P24182
D	-13	HIS	-	expression tag	UNP P24182
D	-12	HIS	-	expression tag	UNP P24182
D	-11	HIS	-	expression tag	UNP P24182
D	-10	HIS	-	expression tag	UNP P24182
D	-9	SER	-	cloning artifact	UNP P24182
D	-8	SER	-	cloning artifact	UNP P24182
D	-7	GLY	-	cloning artifact	UNP P24182
D	-6	LEU	-	cloning artifact	UNP P24182
D	-5	VAL	-	cloning artifact	UNP P24182
D	-4	PRO	-	cloning artifact	UNP P24182
D	-3	ARG	-	cloning artifact	UNP P24182
D	-2	GLY	-	cloning artifact	UNP P24182
D	-1	SER	-	cloning artifact	UNP P24182
D	0	HIS	-	cloning artifact	UNP P24182
D	363	ALA	PHE	engineered mutation	UNP P24182

- Molecule 2 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	272	Total 272	O 272	0	0
2	B	247	Total 247	O 247	0	0
2	C	291	Total 291	O 291	0	0
2	D	261	Total 261	O 261	0	0

3 Residue-property plots ⓘ

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.
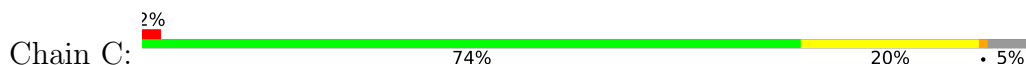
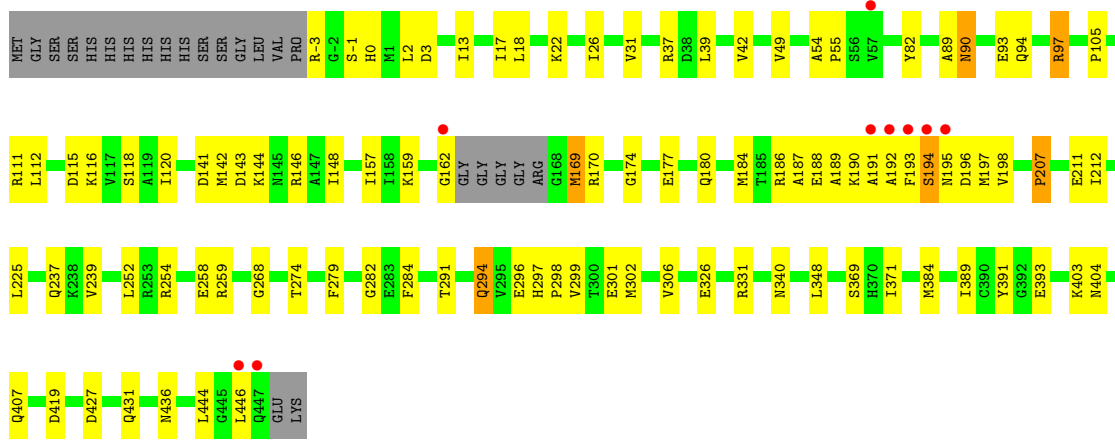- Molecule 1: Biotin carboxylase
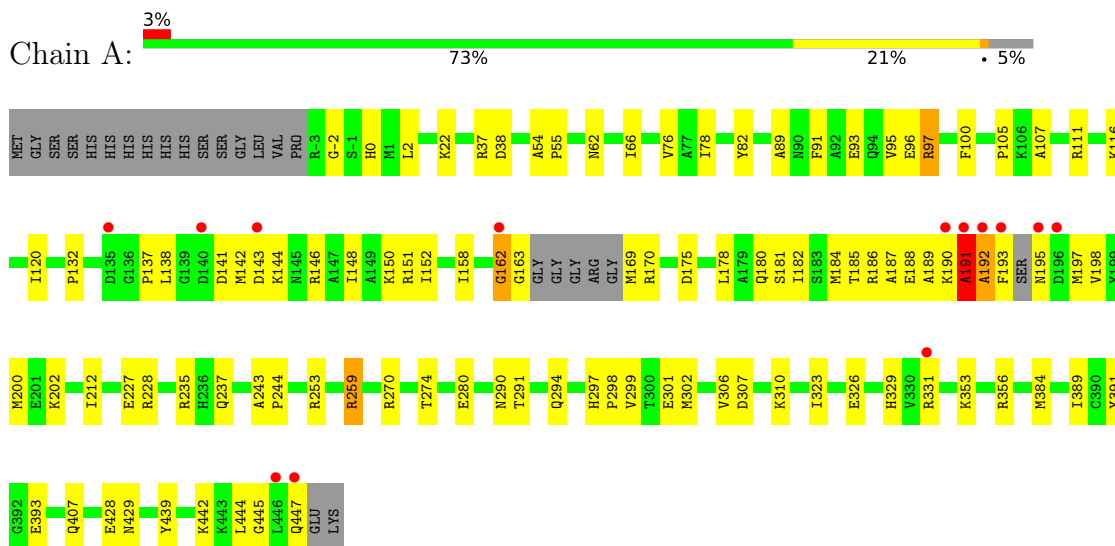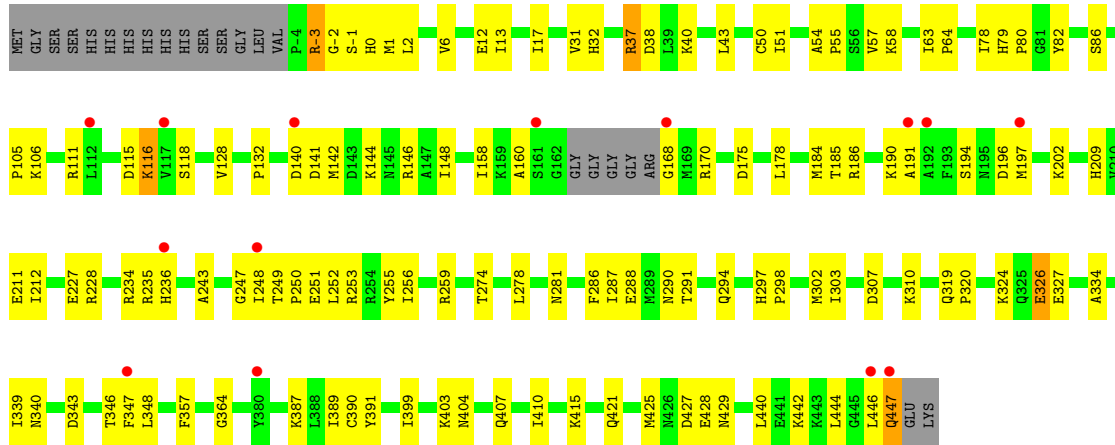
Chain A:

- Molecule 1: Biotin carboxylase

Chain B:

- Molecule 1: Biotin carboxylase

Chain C:

• Molecule 1: Biotin carboxylase

Chain D:
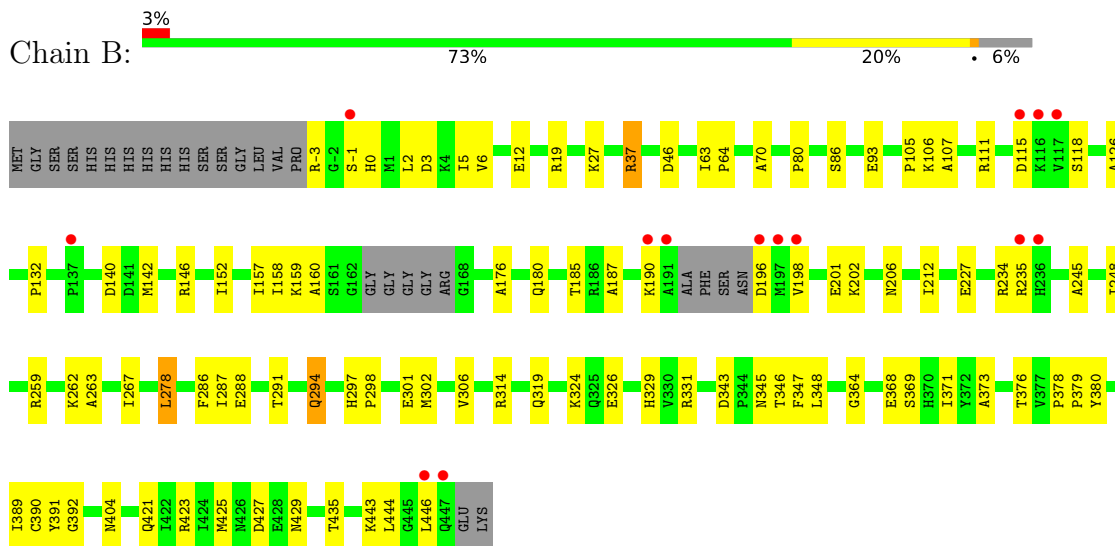
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	62.35Å　　81.50Å　　176.65Å 90.00°　　97.69°　　　90.00°	Depositor
Resolution (Å)	29.75　–　2.20 29.75　–　2.17	Depositor EDS
% Data completeness (in resolution range)	87.8 (29.75-2.20) 85.8 (29.75-2.17)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.42 (at 2.18Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.192　,　0.250 0.192　,　0.249	Depositor DCC
R_{free} test set	6090 reflections (7.55%)	wwPDB-VP
Wilson B-factor (Å^2)	20.3	Xtriage
Anisotropy	0.436	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 51.7	EDS
L-test for twinning[2]	$< |L| > = 0.46, < L^2 > = 0.29$	Xtriage
Estimated twinning fraction	0.054 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	14802	wwPDB-VP
Average B, all atoms (Å^2)	23.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.54% 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	0.38	0/3495	0.62	1/4715 (0.0%)
1	B	0.34	0/3470	0.60	1/4681 (0.0%)
1	C	0.36	0/3502	0.62	3/4726 (0.1%)
1	D	0.33	0/3510	0.59	0/4737
All	All	0.35	0/13977	0.61	5/18859 (0.0%)

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	191	ALA	C-N-CA	-7.04	104.09	121.70
1	A	191	ALA	CA-C-N	-5.51	105.08	117.20
1	C	-1	SER	N-CA-C	-5.19	96.98	111.00
1	B	278	LEU	N-CA-C	-5.09	97.26	111.00
1	C	194	SER	O-C-N	5.04	130.76	122.70

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	3434	0	3459	111	1
1	B	3410	0	3439	88	0
1	C	3440	0	3465	86	1
1	D	3447	0	3472	130	1

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	272	0	0	30	0
2	B	247	0	0	15	0
2	C	291	0	0	23	0
2	D	261	0	0	29	0
All	All	14802	0	13835	408	2

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 15.

The worst 5 of 408 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:C:162:GLY:O	1:C:193:PHE:CE2	1.83	1.30
1:A:162:GLY:O	1:A:193:PHE:CD2	2.01	1.13
1:A:163:GLY:HA2	1:A:193:PHE:CE2	1.97	0.97
1:B:302:MET:HE3	1:B:391:TYR:HB2	1.44	0.96
1:C:162:GLY:O	1:C:193:PHE:CZ	2.18	0.95

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:353:LYS:NZ	1:C:192:ALA:O[1_545]	2.14	0.06
1:D:1:MET:O	1:D:170:ARG:NH2[1_455]	2.19	0.01

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	439/469 (94%)	418 (95%)	17 (4%)	4 (1%)	17	16
1	B	436/469 (93%)	412 (94%)	23 (5%)	1 (0%)	47	55

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	442/469 (94%)	423 (96%)	19 (4%)	0	100	100
1	D	443/469 (94%)	421 (95%)	21 (5%)	1 (0%)	47	55
All	All	1760/1876 (94%)	1674 (95%)	80 (4%)	6 (0%)	41	46

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	191	ALA
1	A	192	ALA
1	A	-2	GLY
1	A	162	GLY
1	B	176	ALA

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	359/377 (95%)	356 (99%)	3 (1%)	81	90
1	B	357/377 (95%)	352 (99%)	5 (1%)	67	80
1	C	360/377 (96%)	349 (97%)	11 (3%)	40	51
1	D	361/377 (96%)	355 (98%)	6 (2%)	60	74
All	All	1437/1508 (95%)	1412 (98%)	25 (2%)	60	74

5 of 25 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	186	ARG
1	C	294	GLN
1	D	447	GLN
1	C	237	GLN
1	C	348	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 41 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	294	GLN
1	D	209	HIS
1	C	319	GLN
1	C	404	ASN
1	D	319	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

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, 95th 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(Å²)	Q<0.9
1	A	445/469 (94%)	-0.07	13 (2%)	51	49	10, 20, 41, 55	2 (0%)
1	B	442/469 (94%)	0.03	14 (3%)	47	45	12, 22, 40, 58	2 (0%)
1	C	446/469 (95%)	-0.13	9 (2%)	65	63	10, 19, 38, 52	2 (0%)
1	D	447/469 (95%)	-0.05	14 (3%)	49	47	9, 20, 39, 52	2 (0%)
All	All	1780/1876 (94%)	-0.05	50 (2%)	53	51	9, 21, 39, 58	8 (0%)

The worst 5 of 50 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	193	PHE	8.8
1	A	193	PHE	8.7
1	A	192	ALA	7.2
1	C	192	ALA	5.6
1	A	191	ALA	5.2

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.