Full wwPDB X-ray Structure Validation Report (i)

Jan 30, 2024 – 12:20 PM EST

PDB ID	:	8T9W
Title	:	Crystal structure of a GH140 apiosidase derived from a lignocellulolytic enriched mangrove metagenome.
Authors	:	Liberato, M.V.; Squina, F.
Deposited on	:	2023-06-26
Resolution	:	2.20 Å(reported)

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

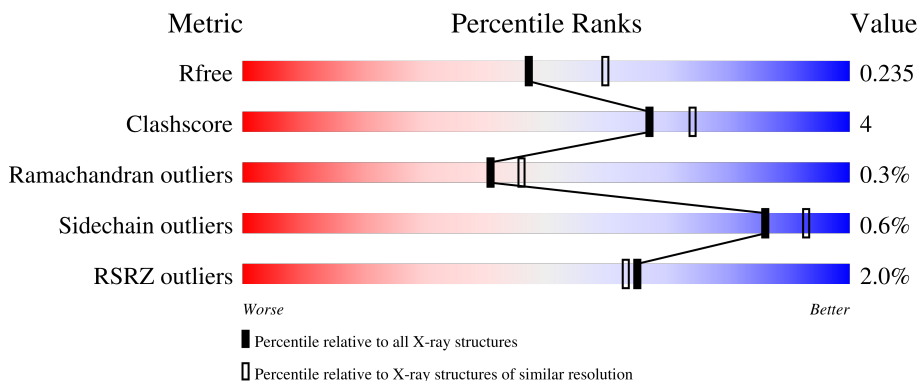|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation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

1 Overall quality at a glance ⓘ

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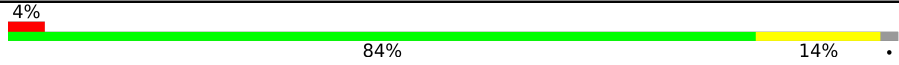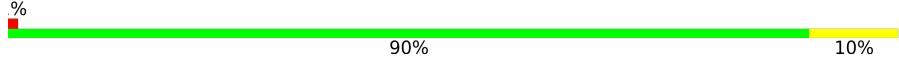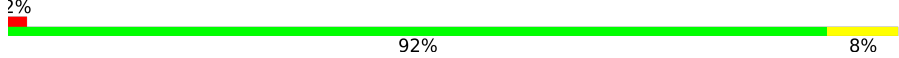*X-RAY DIFFRACTION*

The reported resolution of this entry is 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		0.235
Clashscore		4
Ramachandran outliers		0.3%
Sidechain outliers		0.6%
RSRZ outliers		2.0%

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	434	4% · 84% · 14% ·
1	B	434	% · 89% · 11%
1	C	434	% · 90% · 10%
1	D	434	2% · 92% · 8%

WORLDWIDE
wwPDB
PROTEIN DATA BANK

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 14908 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 Apiosidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	425	Total	C	N	O	S	0	0	0
			3387	2174	578	627	8			
1	B	433	Total	C	N	O	S	0	0	0
			3451	2217	592	634	8			
1	C	434	Total	C	N	O	S	0	0	0
			3497	2247	597	645	8			
1	D	434	Total	C	N	O	S	0	0	0
			3467	2228	594	637	8			

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	238	Total	O	0	0
			238	238		
2	B	210	Total	O	0	0
			210	210		
2	C	353	Total	O	0	0
			353	353		
2	D	305	Total	O	0	0
			305	305		

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.
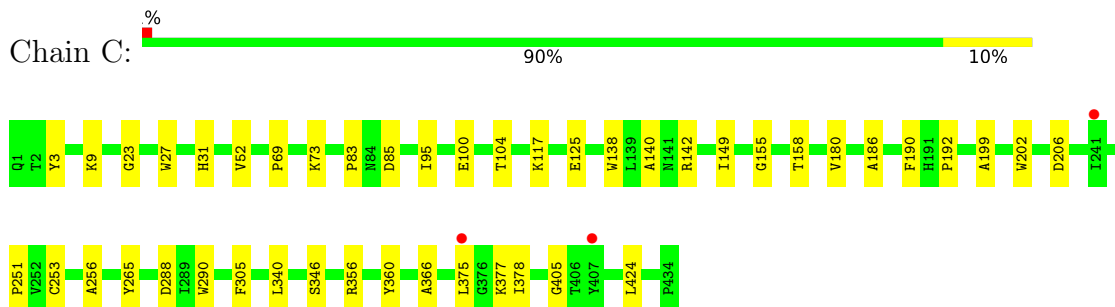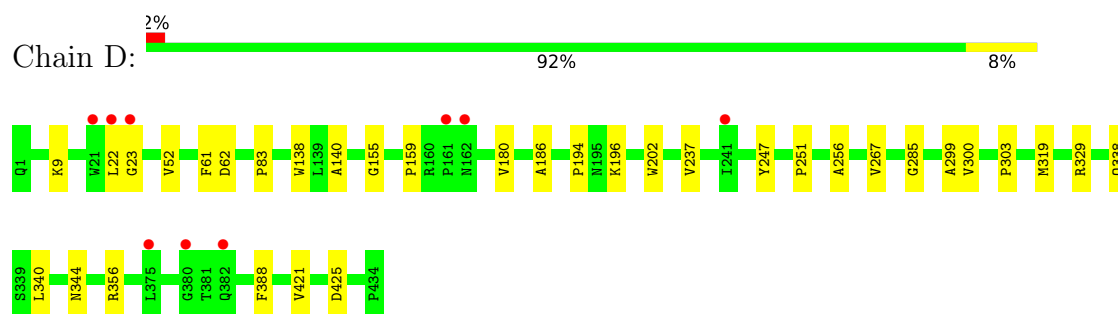
- Molecule 1: Apiosidase

- Molecule 1: Apiosidase

- Molecule 1: Apiosidase

- Molecule 1: Apiosidase

Chain D:
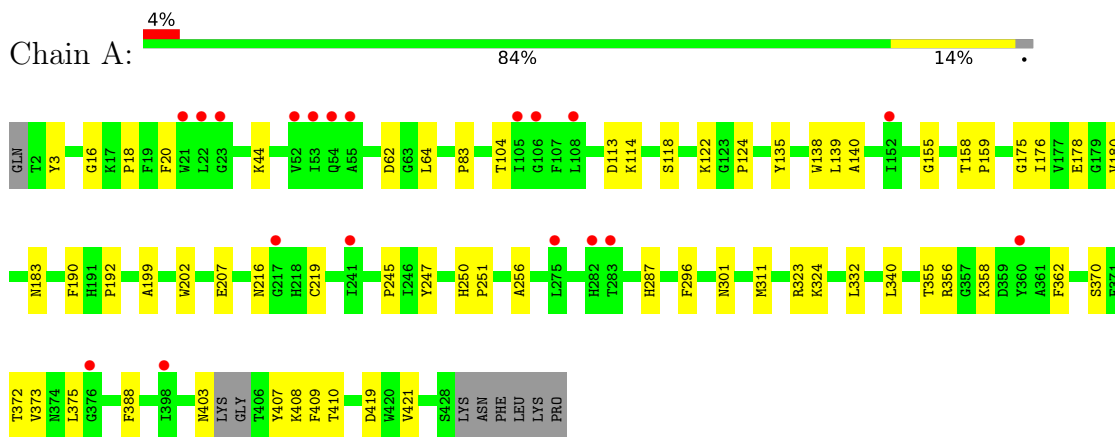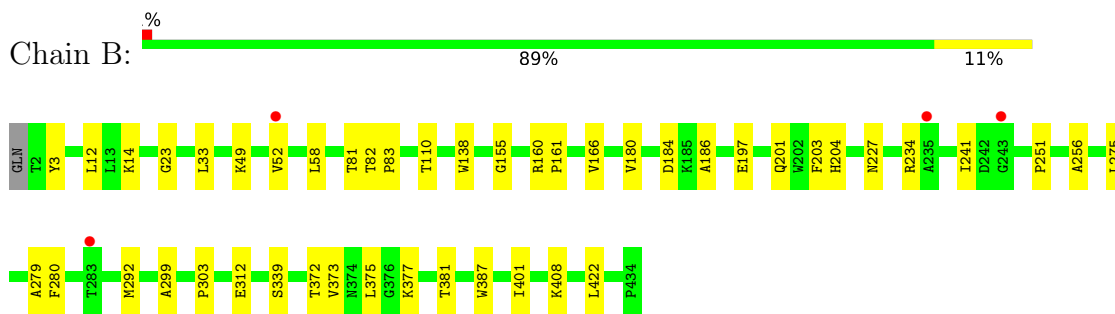
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	112.09Å 113.39Å 158.16Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	37.33 – 2.20 37.33 – 2.20	Depositor EDS
% Data completeness (in resolution range)	99.6 (37.33-2.20) 99.6 (37.33-2.20)	Depositor EDS
R_{merge}	0.20	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.74 (at 2.20Å)	Xtriage
Refinement program	PHENIX 1.20.1-4487	Depositor
R, R_{free}	0.202 , 0.235 0.203 , 0.235	Depositor DCC
R_{free} test set	5228 reflections (5.11%)	wwPDB-VP
Wilson B-factor (Å^2)	26.1	Xtriage
Anisotropy	1.062	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 39.4	EDS
L-test for twinning [2]	$< |L| > = 0.47, < L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.024 for k,h,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	14908	wwPDB-VP
Average B, all atoms (Å^2)	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.07% 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	0.26	0/3492	0.49	0/4756
1	B	0.26	0/3559	0.48	0/4847
1	C	0.27	0/3605	0.49	0/4900
1	D	0.26	0/3575	0.49	0/4864
All	All	0.26	0/14231	0.49	0/19367

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	3387	0	3152	34	0
1	B	3451	0	3235	28	0
1	C	3497	0	3324	26	0
1	D	3467	0	3261	17	0
2	A	238	0	0	1	0
2	B	210	0	0	2	0
2	C	353	0	0	7	0
2	D	305	0	0	2	0
All	All	14908	0	12972	104	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 4.

All (104) 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:375:LEU:O	2:C:501:HOH:O	2.05	0.73
1:C:3:TYR:HH	1:C:104:THR:HG1	1.35	0.71
1:A:18:PRO:HG3	1:A:332:LEU:HD21	1.71	0.70
1:C:117:LYS:HG3	1:C:125:GLU:HG3	1.76	0.68
1:B:186:ALA:O	2:B:501:HOH:O	2.12	0.68
1:A:340:LEU:HD11	1:A:356:ARG:HB3	1.76	0.67
1:A:287:HIS:HE2	1:A:301:ASN:HB3	1.64	0.63
1:D:340:LEU:HD11	1:D:356:ARG:HB3	1.79	0.62
1:C:265:TYR:OH	1:C:366:ALA:O	2.12	0.61
1:D:329:ARG:NH2	1:D:425:ASP:OD2	2.34	0.59
1:B:299:ALA:HB1	1:B:303:PRO:HD2	1.84	0.59
1:B:372:THR:HA	1:B:408:LYS:HA	1.83	0.59
1:C:9:LYS:NZ	2:C:513:HOH:O	2.35	0.59
1:A:375:LEU:HD11	1:A:407:TYR:HB2	1.84	0.59
1:D:196:LYS:NZ	2:D:507:HOH:O	2.33	0.59
1:B:180:VAL:HG21	1:B:186:ALA:HB2	1.87	0.57
1:C:85:ASP:OD1	1:C:142:ARG:NH2	2.29	0.56
1:A:388:PHE:HB3	1:A:421:VAL:HG13	1.87	0.56
1:B:197:GLU:HB2	1:B:201:GLN:HB2	1.88	0.56
1:C:340:LEU:HD11	1:C:356:ARG:HB3	1.86	0.56
1:C:180:VAL:HG21	1:C:186:ALA:HB2	1.87	0.55
1:A:20:PHE:CD1	1:A:323:ARG:HG3	2.41	0.55
1:A:83:PRO:HG2	1:A:138:TRP:CD2	2.44	0.53
1:A:62:ASP:HB2	1:A:122:LYS:HD2	1.91	0.52
1:D:194:PRO:O	2:D:501:HOH:O	2.19	0.52
1:A:183:ASN:HD21	1:A:207:GLU:HG2	1.74	0.52
1:A:287:HIS:NE2	1:A:301:ASN:HB3	2.24	0.52
1:A:207:GLU:OE1	1:A:207:GLU:N	2.37	0.51
1:B:339:SER:OG	1:B:377:LYS:HE2	2.11	0.51
1:D:140:ALA:HB1	1:D:180:VAL:HG13	1.93	0.51
1:B:12:LEU:HD22	1:B:280:PHE:CE1	2.45	0.50
1:C:23:GLY:HA2	1:C:52:VAL:O	2.11	0.50
1:C:3:TYR:OH	1:C:104:THR:OG1	2.12	0.50
1:D:61:PHE:CZ	1:D:300:VAL:HG11	2.47	0.50
1:A:373:VAL:HB	1:A:409:PHE:HE1	1.77	0.49
1:A:3:TYR:OH	1:A:104:THR:OG1	2.16	0.49
1:A:216:ASN:O	1:A:245:PRO:HA	2.12	0.49
1:A:64:LEU:HD12	1:A:124:PRO:HD2	1.94	0.49
1:C:100:GLU:OE2	2:C:502:HOH:O	2.19	0.49
1:D:22:LEU:HD11	1:D:319:MET:HB3	1.95	0.49

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:23:GLY:HA2	1:B:52:VAL:O	2.14	0.48
1:A:370:SER:HB2	1:A:410:THR:HG22	1.95	0.48
1:B:184:ASP:OD1	1:B:184:ASP:N	2.42	0.48
1:C:251:PRO:HB3	1:C:256:ALA:HB2	1.95	0.48
1:A:190:PHE:HB3	1:A:199:ALA:HB1	1.96	0.47
1:A:44:LYS:HG2	1:A:311:MET:HE3	1.97	0.47
1:A:16:GLY:O	2:A:502:HOH:O	2.20	0.46
1:B:381:THR:N	2:B:526:HOH:O	2.47	0.46
1:D:251:PRO:HB3	1:D:256:ALA:HB2	1.97	0.46
1:C:95:ILE:HD13	1:C:149:ILE:HD11	1.97	0.46
1:A:140:ALA:HB1	1:A:180:VAL:HG13	1.99	0.45
1:A:247:TYR:HB2	1:A:250:HIS:HB2	1.98	0.45
1:A:296:PHE:CE2	1:B:312:GLU:HB2	2.50	0.45
1:A:140:ALA:HB2	1:A:176:ILE:HA	1.97	0.45
1:A:358:LYS:HE3	1:A:358:LYS:HB2	1.54	0.45
1:B:83:PRO:HG2	1:B:138:TRP:CD2	2.52	0.45
1:D:299:ALA:HB1	1:D:303:PRO:HD2	1.99	0.45
1:D:23:GLY:HA2	1:D:52:VAL:O	2.15	0.45
1:C:158:THR:HG23	1:C:202:TRP:CZ2	2.52	0.45
1:D:338:GLN:HB3	1:D:344:ASN:HB2	1.99	0.45
1:C:377:LYS:N	2:C:501:HOH:O	2.50	0.45
1:B:375:LEU:HD21	1:B:401:ILE:HD12	1.99	0.44
1:D:83:PRO:HG2	1:D:138:TRP:CD2	2.52	0.44
1:A:159:PRO:HD2	1:A:202:TRP:CZ2	2.52	0.44
1:C:378:ILE:HG12	2:C:501:HOH:O	2.16	0.44
1:D:9:LYS:HB2	1:D:237:VAL:HG12	1.99	0.44
1:B:3:TYR:CE1	1:B:14:LYS:HD2	2.52	0.44
1:C:190:PHE:HB3	1:C:199:ALA:HB1	2.00	0.44
1:B:81:THR:HG23	1:B:82:THR:HG23	2.00	0.44
1:B:373:VAL:HG12	1:B:375:LEU:HD12	1.99	0.44
1:B:166:VAL:HG13	1:B:203:PHE:HE1	1.83	0.44
1:B:241:ILE:HB	1:B:280:PHE:CE2	2.53	0.44
1:C:73:LYS:NZ	2:C:535:HOH:O	2.51	0.44
1:C:140:ALA:HB1	1:C:180:VAL:HG13	1.99	0.44
1:A:355:THR:HG22	1:A:362:PHE:HB2	2.00	0.43
1:A:419:ASP:OD1	1:A:419:ASP:N	2.49	0.43
1:B:49:LYS:HD3	1:B:49:LYS:HA	1.76	0.43
1:C:9:LYS:HG2	2:C:573:HOH:O	2.18	0.43
1:D:180:VAL:HG21	1:D:186:ALA:HB2	2.01	0.43
1:A:324:LYS:HB3	1:A:324:LYS:HE2	1.89	0.43
1:A:113:ASP:OD1	1:A:114:LYS:NZ	2.50	0.43

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:58:LEU:HD12	1:B:110:THR:HB	2.00	0.42
1:A:251:PRO:HB3	1:A:256:ALA:HB2	2.01	0.42
1:D:159:PRO:HD2	1:D:202:TRP:CZ2	2.53	0.42
1:B:275:LEU:HA	1:B:279:ALA:HB3	2.01	0.42
1:C:375:LEU:HB2	1:C:405:GLY:O	2.19	0.42
1:A:158:THR:HG23	1:A:202:TRP:CZ2	2.55	0.42
1:B:3:TYR:HE1	1:B:14:LYS:HD2	1.85	0.42
1:A:135:TYR:CZ	1:A:139:LEU:HD11	2.55	0.41
1:B:33:LEU:HG	1:B:292:MET:HG3	2.02	0.41
1:D:388:PHE:HB3	1:D:421:VAL:HB	2.01	0.41
1:B:251:PRO:HB3	1:B:256:ALA:HB2	2.03	0.41
1:C:288:ASP:HB3	1:C:305:PHE:O	2.21	0.41
1:B:166:VAL:HG13	1:B:203:PHE:CE1	2.56	0.41
1:C:31:HIS:O	1:C:69:PRO:HD3	2.21	0.41
1:C:83:PRO:HG2	1:C:138:TRP:CD2	2.56	0.41
1:A:372:THR:HA	1:A:408:LYS:HA	2.03	0.41
1:C:27:TRP:CD1	1:C:290:TRP:HB2	2.56	0.40
1:D:247:TYR:CD1	1:D:285:GLY:HA3	2.56	0.40
1:A:175:GLY:HA2	1:A:178:GLU:HB3	2.03	0.40
1:B:204:HIS:CD2	1:B:234:ARG:HD3	2.55	0.40
1:B:387:TRP:CZ3	1:B:422:LEU:HB2	2.57	0.40
1:B:160:ARG:HA	1:B:161:PRO:HD3	1.94	0.40
1:C:360:TYR:HA	1:C:424:LEU:O	2.22	0.40

There are no symmetry-related clashes.

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	421/434 (97%)	407 (97%)	12 (3%)	2 (0%)	29	31
1	B	431/434 (99%)	414 (96%)	16 (4%)	1 (0%)	47	55

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	432/434 (100%)	416 (96%)	14 (3%)	2 (0%)	29	31
1	D	432/434 (100%)	412 (95%)	19 (4%)	1 (0%)	47	55
All	All	1716/1736 (99%)	1649 (96%)	61 (4%)	6 (0%)	41	46

All (6) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	155	GLY
1	B	155	GLY
1	C	155	GLY
1	D	155	GLY
1	C	192	PRO
1	A	192	PRO

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	341/365 (93%)	338 (99%)	3 (1%)	78	88
1	B	349/365 (96%)	348 (100%)	1 (0%)	92	97
1	C	360/365 (99%)	357 (99%)	3 (1%)	81	90
1	D	350/365 (96%)	348 (99%)	2 (1%)	86	93
All	All	1400/1460 (96%)	1391 (99%)	9 (1%)	86	93

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	118	SER
1	A	219	CYS
1	A	403	ASN
1	B	227	ASN
1	C	206	ASP
1	C	253	CYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	346	SER
1	D	62	ASP
1	D	267	VAL

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

Mol	Chain	Res	Type
1	A	403	ASN
1	B	227	ASN
1	B	274	ASN
1	D	1	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	425/434 (97%)	0.10	19 (4%)	33	32	27, 41, 67, 82	0
1	B	433/434 (99%)	-0.02	4 (0%)	84	83	25, 45, 60, 71	0
1	C	434/434 (100%)	-0.27	3 (0%)	87	86	18, 29, 45, 58	0
1	D	434/434 (100%)	-0.11	9 (2%)	63	61	19, 32, 58, 78	0
All	All	1726/1736 (99%)	-0.07	35 (2%)	65	63	18, 37, 60, 82	0

All (35) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	283	THR	4.1
1	A	53	ILE	3.6
1	D	161	PRO	3.6
1	A	108	LEU	3.1
1	D	162	ASN	3.1
1	A	55	ALA	3.1
1	B	235	ALA	3.1
1	A	52	VAL	2.9
1	D	375	LEU	2.9
1	A	241	ILE	2.8
1	A	376	GLY	2.7
1	C	407	TYR	2.7
1	D	22	LEU	2.6
1	A	21	TRP	2.6
1	B	52	VAL	2.5
1	A	398	ILE	2.5
1	A	23	GLY	2.5
1	B	283	THR	2.5
1	A	152	ILE	2.5
1	C	241	ILE	2.4
1	D	241	ILE	2.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	22	LEU	2.4
1	A	275	LEU	2.4
1	D	23	GLY	2.3
1	D	380	GLY	2.2
1	C	375	LEU	2.2
1	A	105	ILE	2.2
1	A	106	GLY	2.2
1	A	360	TYR	2.1
1	A	282	HIS	2.1
1	D	21	TRP	2.1
1	A	217	GLY	2.1
1	D	382	GLN	2.0
1	A	54	GLN	2.0
1	B	243	GLY	2.0

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 ⓘ

There are no such residues in this entry.