wwPDB X-ray Structure Validation Summary Report (i)

Jan 21, 2025 – 02:13 PM JST

PDB ID	:	8JZQ
Title	:	Crystal structure of Panax quinquefolius Pq3-O-UGT2
Authors	:	Mei, K.; Ji, Q.; Liu, Y.
Deposited on	:	2023-07-06
Resolution	:	2.89 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 (i)) were used in the production of this report:

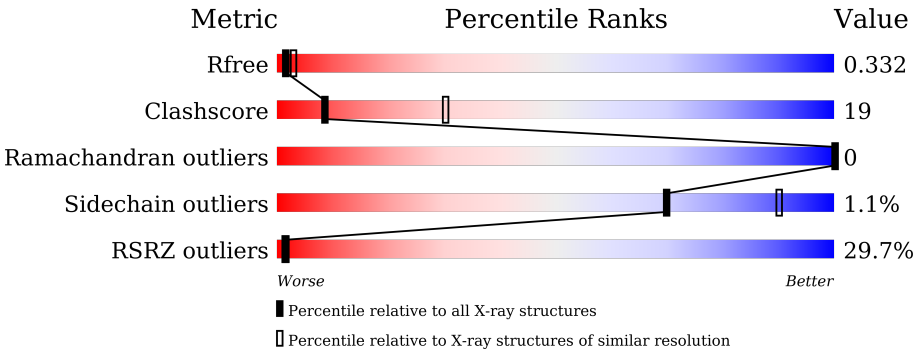
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.21
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.004 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.40

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 2.89 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}	164625	2335 (2.90-2.90)
Clashscore	180529	2564 (2.90-2.90)
Ramachandran outliers	177936	2514 (2.90-2.90)
Sidechain outliers	177891	2516 (2.90-2.90)
RSRZ outliers	164620	2337 (2.90-2.90)

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	455	15% 64% 18% 17%
1	B	455	26% 64% 19% 17%
1	C	455	28% 64% 19% 17%
1	D	455	31% 64% 19% 17%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 11872 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 Glycosyltransferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	378	Total	C	N	O	S	0	0	0
			2968	1902	500	554	12			
1	B	378	Total	C	N	O	S	0	0	0
			2968	1902	500	554	12			
1	C	378	Total	C	N	O	S	0	0	0
			2968	1902	500	554	12			
1	D	378	Total	C	N	O	S	0	0	0
			2968	1902	500	554	12			

There are 52 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	443	GLU	-	expression tag	UNP A0A0M4ME80
A	444	ASN	-	expression tag	UNP A0A0M4ME80
A	445	LEU	-	expression tag	UNP A0A0M4ME80
A	446	TYR	-	expression tag	UNP A0A0M4ME80
A	447	PHE	-	expression tag	UNP A0A0M4ME80
A	448	GLN	-	expression tag	UNP A0A0M4ME80
A	449	GLY	-	expression tag	UNP A0A0M4ME80
A	450	HIS	-	expression tag	UNP A0A0M4ME80
A	451	HIS	-	expression tag	UNP A0A0M4ME80
A	452	HIS	-	expression tag	UNP A0A0M4ME80
A	453	HIS	-	expression tag	UNP A0A0M4ME80
A	454	HIS	-	expression tag	UNP A0A0M4ME80
A	455	HIS	-	expression tag	UNP A0A0M4ME80
B	443	GLU	-	expression tag	UNP A0A0M4ME80
B	444	ASN	-	expression tag	UNP A0A0M4ME80
B	445	LEU	-	expression tag	UNP A0A0M4ME80
B	446	TYR	-	expression tag	UNP A0A0M4ME80
B	447	PHE	-	expression tag	UNP A0A0M4ME80
B	448	GLN	-	expression tag	UNP A0A0M4ME80
B	449	GLY	-	expression tag	UNP A0A0M4ME80
B	450	HIS	-	expression tag	UNP A0A0M4ME80

Continued on next page...

Continued from previous page...

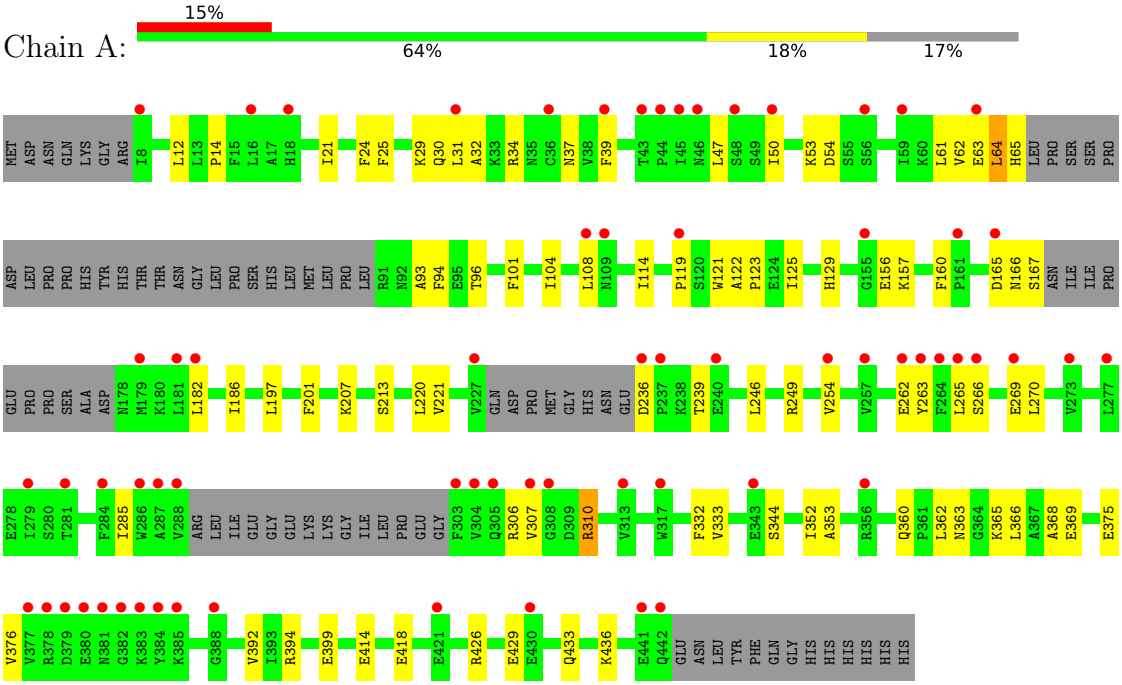
Chain	Residue	Modelled	Actual	Comment	Reference
B	451	HIS	-	expression tag	UNP A0A0M4ME80
B	452	HIS	-	expression tag	UNP A0A0M4ME80
B	453	HIS	-	expression tag	UNP A0A0M4ME80
B	454	HIS	-	expression tag	UNP A0A0M4ME80
B	455	HIS	-	expression tag	UNP A0A0M4ME80
C	443	GLU	-	expression tag	UNP A0A0M4ME80
C	444	ASN	-	expression tag	UNP A0A0M4ME80
C	445	LEU	-	expression tag	UNP A0A0M4ME80
C	446	TYR	-	expression tag	UNP A0A0M4ME80
C	447	PHE	-	expression tag	UNP A0A0M4ME80
C	448	GLN	-	expression tag	UNP A0A0M4ME80
C	449	GLY	-	expression tag	UNP A0A0M4ME80
C	450	HIS	-	expression tag	UNP A0A0M4ME80
C	451	HIS	-	expression tag	UNP A0A0M4ME80
C	452	HIS	-	expression tag	UNP A0A0M4ME80
C	453	HIS	-	expression tag	UNP A0A0M4ME80
C	454	HIS	-	expression tag	UNP A0A0M4ME80
C	455	HIS	-	expression tag	UNP A0A0M4ME80
D	443	GLU	-	expression tag	UNP A0A0M4ME80
D	444	ASN	-	expression tag	UNP A0A0M4ME80
D	445	LEU	-	expression tag	UNP A0A0M4ME80
D	446	TYR	-	expression tag	UNP A0A0M4ME80
D	447	PHE	-	expression tag	UNP A0A0M4ME80
D	448	GLN	-	expression tag	UNP A0A0M4ME80
D	449	GLY	-	expression tag	UNP A0A0M4ME80
D	450	HIS	-	expression tag	UNP A0A0M4ME80
D	451	HIS	-	expression tag	UNP A0A0M4ME80
D	452	HIS	-	expression tag	UNP A0A0M4ME80
D	453	HIS	-	expression tag	UNP A0A0M4ME80
D	454	HIS	-	expression tag	UNP A0A0M4ME80
D	455	HIS	-	expression tag	UNP A0A0M4ME80

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.

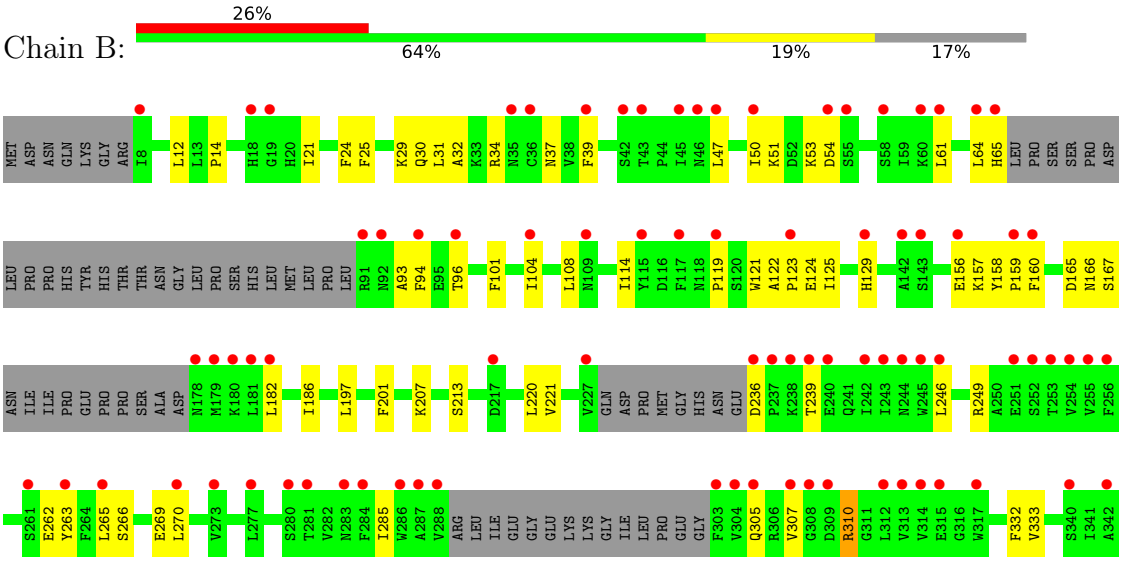
- Molecule 1: Glycosyltransferase

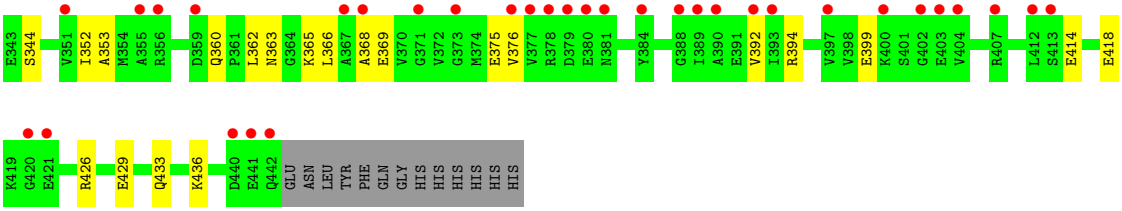
Chain A:



- Molecule 1: Glycosyltransferase

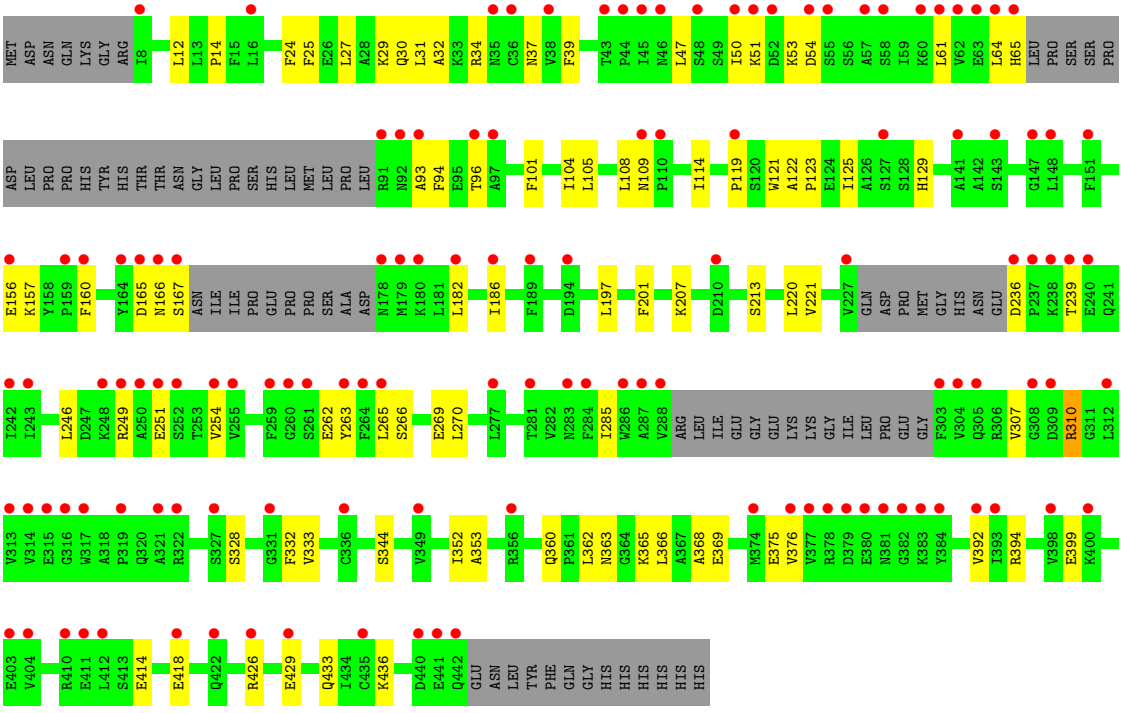
Chain B:





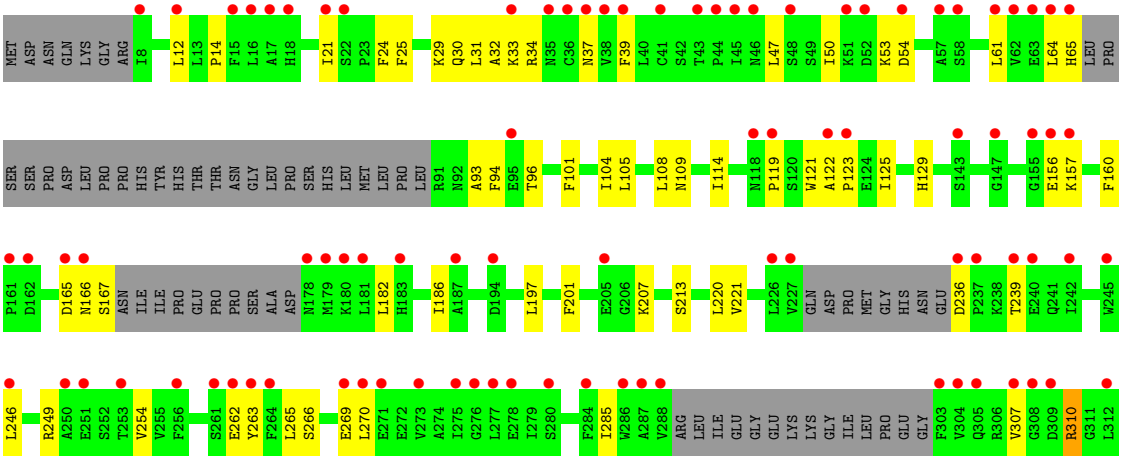
• Molecule 1: Glycosyltransferase

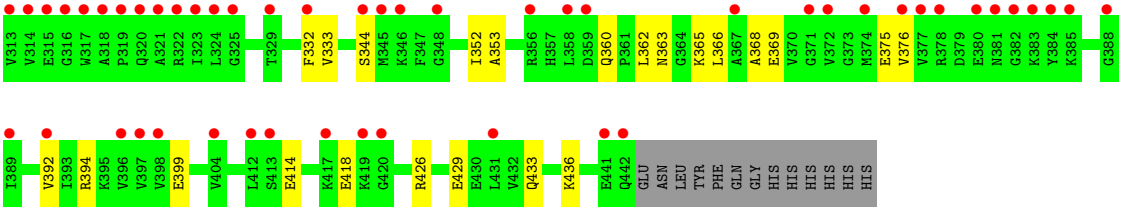
Chain C:



• Molecule 1: Glycosyltransferase

Chain D:




V313
V314
E315
G316
W317
A318
P319
Q320
A321
R322
I323
L324
G325
T329
F332
V333
S344
M345
K346
F347
G348
I352
A353
R356
H357
L358
D359
Q360
P361
L362
N363
G364
K365
L366
A367
A368
E369
V370
G371
V372
G373
M374
E375
V376
V377
R378
D379
E380
N381
G382
K383
Y384
K385
G388
I389
V392
I393
R394
K395
V396
V397
V398
E399
V404
L412
S413
E414
K417
E418
K419
G420
R426
E429
E430
L431
V432
Q433
K436
E441
Q442
GLU
ASN
LEU
TYR
PHE
GLN
GLY
HIS
HIS
HIS
HIS
HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	46.99Å 80.89Å 126.38Å 93.96° 90.00° 93.75°	Depositor
Resolution (Å)	37.70 – 2.89 37.70 – 2.89	Depositor EDS
% Data completeness (in resolution range)	94.3 (37.70-2.89) 94.5 (37.70-2.89)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.43 (at 2.91Å)	Xtriage
Refinement program	PHENIX (1.19_4092: ???)	Depositor
R, R_{free}	0.307 , 0.333 0.306 , 0.332	Depositor DCC
R_{free} test set	2053 reflections (4.92%)	wwPDB-VP
Wilson B-factor (Å^2)	62.6	Xtriage
Anisotropy	0.329	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 50.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.83	EDS
Total number of atoms	11872	wwPDB-VP
Average B, all atoms (Å^2)	71.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.33% 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	0.42	0/3027	0.59	0/4086
1	B	0.42	0/3027	0.58	0/4086
1	C	0.42	0/3027	0.58	0/4086
1	D	0.42	0/3027	0.58	0/4086
All	All	0.42	0/12108	0.58	0/16344

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	2968	0	2984	111	7
1	B	2968	0	2984	107	10
1	C	2968	0	2984	113	6
1	D	2968	0	2983	111	8
All	All	11872	0	11935	442	27

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 19.

The worst 5 of 442 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:265:LEU:HD11	1:B:269:GLU:CB	1.07	1.54
1:C:265:LEU:HD11	1:C:269:GLU:CB	1.07	1.54
1:B:265:LEU:CD1	1:B:269:GLU:CB	1.87	1.52
1:D:265:LEU:CD1	1:D:269:GLU:CB	1.87	1.52
1:D:265:LEU:HD11	1:D:269:GLU:CB	1.07	1.51

The worst 5 of 27 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:D:64:LEU:CD1	1:D:156:GLU:OE1[1_455]	1.09	1.11
1:B:64:LEU:CD1	1:B:156:GLU:OE1[1_455]	1.10	1.10
1:A:64:LEU:CD2	1:A:156:GLU:OE1[1_655]	1.17	1.03
1:A:64:LEU:CB	1:A:156:GLU:OE2[1_655]	1.36	0.84
1:B:414:GLU:OE2	1:D:33:LYS:NZ[1_565]	1.39	0.81

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	368/455 (81%)	339 (92%)	29 (8%)	0	100	100
1	B	368/455 (81%)	337 (92%)	31 (8%)	0	100	100
1	C	368/455 (81%)	337 (92%)	31 (8%)	0	100	100
1	D	368/455 (81%)	337 (92%)	31 (8%)	0	100	100
All	All	1472/1820 (81%)	1350 (92%)	122 (8%)	0	100	100

There are no Ramachandran outliers to report.

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	327/396 (83%)	322 (98%)	5 (2%)	60 85
1	B	327/396 (83%)	324 (99%)	3 (1%)	75 92
1	C	327/396 (83%)	324 (99%)	3 (1%)	75 92
1	D	327/396 (83%)	324 (99%)	3 (1%)	75 92
All	All	1308/1584 (83%)	1294 (99%)	14 (1%)	70 90

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

Mol	Chain	Res	Type
1	B	310	ARG
1	C	94	PHE
1	D	310	ARG
1	D	94	PHE
1	D	201	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 oligosaccharides 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	378/455 (83%)	1.12	67 (17%)	4	4	30, 57, 104, 143	0
1	B	378/455 (83%)	1.57	117 (30%)	1	1	33, 63, 116, 156	0
1	C	378/455 (83%)	1.66	126 (33%)	1	1	30, 65, 112, 169	0
1	D	378/455 (83%)	1.81	139 (36%)	1	1	37, 73, 129, 157	0
All	All	1512/1820 (83%)	1.54	449 (29%)	1	1	30, 64, 118, 169	0

The worst 5 of 449 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	251	GLU	7.4
1	D	313	VAL	7.3
1	C	237	PRO	7.2
1	D	45	ILE	6.9
1	C	263	TYR	6.6

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.