Full wwPDB X-ray Structure Validation Report ⓘ

May 14, 2020 – 06:09 am BST

PDB ID	:	5M7R
Title	:	Structure of human O-GlcNAc hydrolase
Authors	:	Roth, C.; Chan, S.; Offen, W.A.; Hemsworth, G.R.; Willems, L.I.; King, D.; Varghese, V.; Britton, R.; Vocadlo, D.J.; Davies, G.J.
Deposited on	:	2016-10-28
Resolution	:	2.35 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 ⓘ symbol.

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

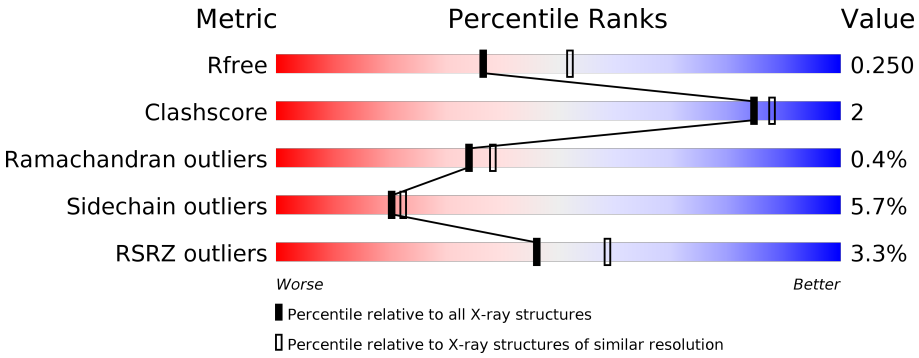
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance

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

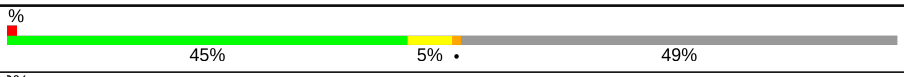
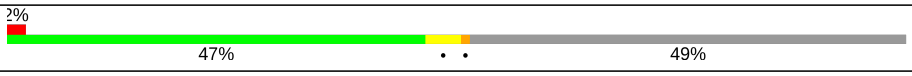
The reported resolution of this entry is 2.35 Å.

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	1164 (2.36-2.36)
Clashscore	141614	1232 (2.36-2.36)
Ramachandran outliers	138981	1211 (2.36-2.36)
Sidechain outliers	138945	1212 (2.36-2.36)
RSRZ outliers	127900	1150 (2.36-2.36)

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 >=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	916	
1	B	916	

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 7704 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 Protein O-GlcNAcase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	466	Total	C	N	O	S	0	0	0
			3813	2468	628	693	24			
1	B	470	Total	C	N	O	S	0	0	0
			3839	2481	634	699	25			

- Molecule 2 is water.

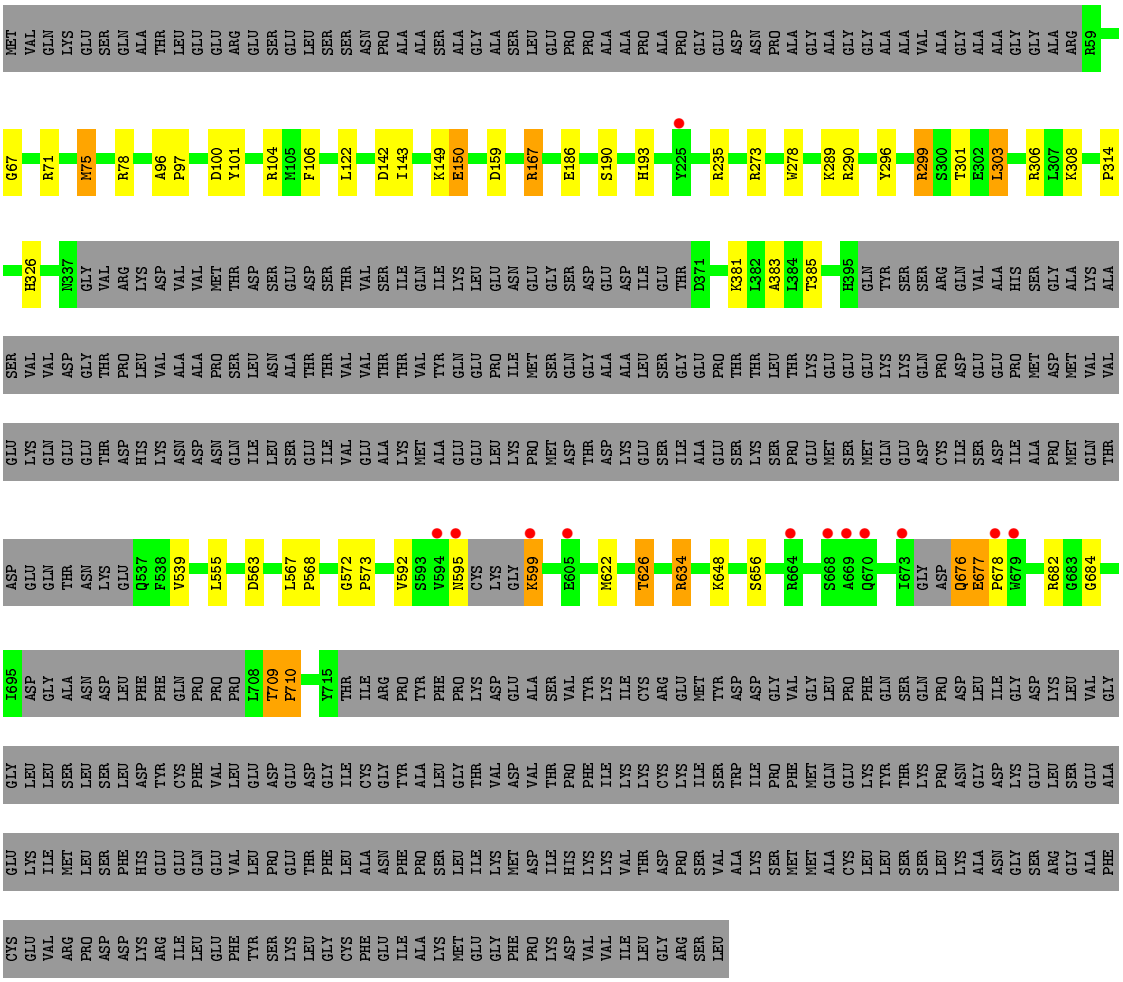
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	21	Total	O	0	0
			21	21		
2	B	31	Total	O	0	0
			31	31		

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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.

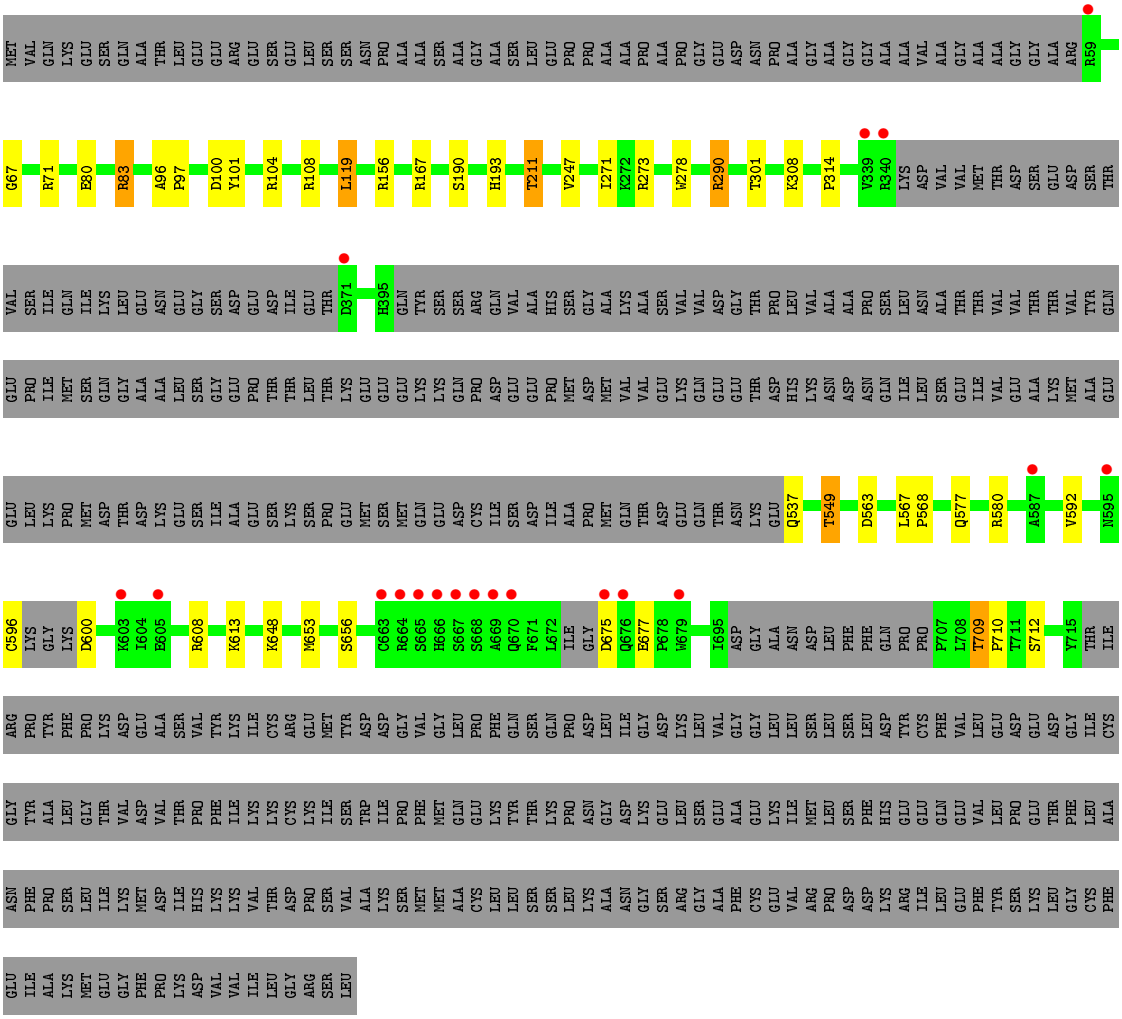
- Molecule 1: Protein O-GlcNAcase

Chain A:



- Molecule 1: Protein O-GlcNAcase

Chain B:


MET VAL GLN LYS GLU SER GLN ALA THR LEU GLU GLU GLU ARG GLU SER GLU LEU SER SER ASN PRO ALA ALA SER ALA GLY ALA SER LEU GLU PRO PRO ALA ALA PRO ALA PRO GLY GLU ASP ASN PRO ALA GLY ALA GLY GLY GLY ALA ALA VAL ALA GLY ALA ALA ALA GLY GLY ALA ARG R59
G67 R71 E80 R83 A96 P97 D100 Y101 R104 R108 L119 R156 R167 S190 H193 T211 V247 I271 R272 R273 W278 R290 T301 K308 P314 V339 R340 LYS ASP VAL VAL MET THR ASP SER GLU ASP SER THR
VAL SER ILE GLN ILE LYS LEU GLU ASN GLU GLY SER ASP GLU ASP ILE GLU THR D371 H395 GLN TYR SER SER ARG GLN VAL ALA HIS SER GLY ALA LYS ALA SER VAL VAL ASP GLY THR PRO LEU VAL ALA ALA PRO SER LEU ASN ALA THR THR VAL VAL THR THR VAL TYR GLN
GLU PRO ILE MET SER GLN GLY ALA ALA LEU SER GLY GLU PRO THR THR LEU THR LYS GLU GLU GLU LYS LYS GLN PRO ASP GLU GLU PRO MET ASP MET VAL VAL GLU LYS GLN GLU GLU GLU THR ASP HIS LYS ASN ASP ASN GLN ILE LEU SER GLU ILE VAL GLU ALA LYS MET ALA GLU
GLU LEU LYS PRO MET ASP THR ASP LYS GLU SER LYS SER PRO GLU MET SER MET GLN GLU ASP CYS ILE SER ASP ILE ALA PRO MET GLN THR ASP GLU GLN THR ASN LYS GLU Q537 T549 D563 L567 P568 Q577 R580 A587 V592 N595
C596 LYS GLY LYS D600 R603 I604 E605 R608 K613 K648 M653 S656 C663 R664 S665 H666 S667 S668 A669 Q670 F671 L672 ILE GLY D675 Q676 E677 P678 W679 I695 ASP GLY ALA ASN ASP LEU PHE PHE GLN PRO PRO F707 L708 T709 P710 T711 S712 Y715 THR ILE
ARG PRO TYR PHE PRO LYS ASP GLU ALA SER VAL TYR LYS ILE CYS ARG GLU MET TYR ASP ASP GLY VAL GLY LEU PRO PHE GLN SER GLN PRO ASP LEU ILE GLY ASP LYS LEU VAL GLY GLY LEU LEU SER LEU SER LEU ASP TYR CYS PHE VAL LEU GLU ASP GLU ASP GLY ILE CYS
GLY TYR ALA LEU GLY THR VAL ASP VAL THR PRO PHE ILE LYS LYS CYS LYS ILE SER TRP ILE PRO PHE MET GLN GLU LYS TYR THR LYS PRO ASN GLY ASP LYS GLU LEU SER GLU ALA GLU LYS ILE MET LEU SER LEU SER PHE HIS GLU GLU GLN GLU VAL LEU LEU PRO GLU THR PHE LEU ALA
ASN PHE PRO SER LEU ILE LYS MET ASP ILE HIS LYS LYS VAL THR ASP PRO SER VAL ALA LYS SER MET MET ALA CYS LEU LEU SER SER LEU LYS ALA ASN GLY SER ARG GLY ALA PHE CYS GLU VAL ARG PRO ASP ASP LYS ARG ILE LEU GLU GLU PHE TYR SER LYS LEU GLY CYS PHE
GLU ILE ALA LYS MET GLU GLY PHE PRO LYS ASP VAL VAL ILE LEU GLY ARG SER LEU

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	102.23Å 102.23Å 285.49Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	96.25 – 2.35 96.25 – 2.35	Depositor EDS
% Data completeness (in resolution range)	100.0 (96.25-2.35) 100.0 (96.25-2.35)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.51 (at 2.34Å)	Xtriage
Refinement program	REFMAC 5.8.0155, REFMAC 5.8.0155	Depositor
R, R_{free}	0.216 , 0.249 0.219 , 0.250	Depositor DCC
R_{free} test set	1909 reflections (2.98%)	wwPDB-VP
Wilson B-factor (Å^2)	59.3	Xtriage
Anisotropy	0.630	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 39.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	7704	wwPDB-VP
Average B, all atoms (Å^2)	80.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.08% 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	0.85	1/3915 (0.0%)	0.97	15/5302 (0.3%)
1	B	0.66	0/3942	0.84	8/5339 (0.1%)
All	All	0.76	1/7857 (0.0%)	0.91	23/10641 (0.2%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	677	GLU	CG-CD	6.59	1.61	1.51

All (23) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	634	ARG	NE-CZ-NH1	9.74	125.17	120.30
1	A	634	ARG	NE-CZ-NH2	-7.99	116.30	120.30
1	A	167	ARG	CG-CD-NE	7.63	127.82	111.80
1	B	290	ARG	NE-CZ-NH2	-7.62	116.49	120.30
1	A	167	ARG	NE-CZ-NH1	7.53	124.06	120.30
1	A	167	ARG	CB-CG-CD	7.23	130.39	111.60
1	B	108	ARG	NE-CZ-NH1	6.99	123.79	120.30
1	A	290	ARG	NE-CZ-NH2	-6.82	116.89	120.30
1	A	303	LEU	CA-CB-CG	6.72	130.76	115.30
1	A	159	ASP	CB-CA-C	6.49	123.38	110.40
1	B	104	ARG	NE-CZ-NH1	5.98	123.29	120.30
1	A	677	GLU	OE1-CD-OE2	-5.93	116.18	123.30
1	A	71	ARG	NE-CZ-NH1	5.92	123.26	120.30
1	A	563	ASP	CB-CG-OD1	5.90	123.61	118.30
1	A	306	ARG	NE-CZ-NH1	5.62	123.11	120.30
1	B	71	ARG	NE-CZ-NH1	5.45	123.02	120.30
1	A	104	ARG	NE-CZ-NH2	-5.44	117.58	120.30
1	B	119	LEU	CA-CB-CG	5.40	127.71	115.30
1	A	626	THR	N-CA-CB	5.35	120.47	110.30
1	A	104	ARG	NE-CZ-NH1	5.33	122.97	120.30
1	B	580	ARG	NE-CZ-NH1	5.16	122.88	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	83	ARG	NE-CZ-NH1	5.16	122.88	120.30
1	B	563	ASP	CB-CG-OD1	5.14	122.93	118.30

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	3813	0	3741	22	0
1	B	3839	0	3759	15	0
2	A	21	0	0	0	0
2	B	31	0	0	1	0
All	All	7704	0	7500	35	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 2.

All (35) 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:A:75:MET:HE1	1:A:122:LEU:HD22	1.33	1.06
1:B:596:CYS:HG	1:B:600:ASP:N	1.61	0.98
1:A:75:MET:CE	1:A:122:LEU:HD22	2.16	0.73
1:A:75:MET:HA	1:A:75:MET:HE3	1.77	0.67
1:A:595:ASN:O	1:A:599:LYS:N	2.32	0.62
1:A:709:THR:HG22	1:A:710:PRO:HD2	1.85	0.57
1:A:709:THR:HG22	1:A:710:PRO:CD	2.37	0.55
1:A:143:ILE:HG13	1:A:150:GLU:HG2	1.88	0.54
1:B:596:CYS:SG	1:B:600:ASP:N	2.74	0.54
1:B:592:VAL:HG12	1:B:592:VAL:O	2.08	0.54
1:A:676:GLN:HE21	1:A:676:GLN:HA	1.71	0.53
1:A:190:SER:OG	1:A:193:HIS:HD2	1.92	0.52
1:A:106:PHE:CE1	1:B:549:THR:HG21	2.45	0.51
1:A:592:VAL:HG12	1:A:592:VAL:O	2.10	0.51

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:190:SER:OG	1:B:193:HIS:HD2	1.95	0.50
1:B:709:THR:HG22	1:B:710:PRO:CD	2.42	0.50
1:B:80:GLU:OE2	1:B:83:ARG:NH1	2.45	0.49
1:A:143:ILE:HG13	1:A:150:GLU:CG	2.43	0.49
1:A:67:GLY:HA2	1:A:96:ALA:O	2.13	0.48
1:A:684:GLY:CA	1:B:290:ARG:HD3	2.44	0.48
1:A:385:THR:HG22	1:A:555:LEU:HD22	1.95	0.47
1:B:709:THR:HG22	1:B:710:PRO:HD2	1.97	0.47
1:B:67:GLY:HA2	1:B:96:ALA:O	2.16	0.46
1:A:97:PRO:HG2	1:A:100:ASP:HB2	1.97	0.46
1:A:567:LEU:HB3	1:A:568:PRO:HD2	1.99	0.45
1:B:567:LEU:HB3	1:B:568:PRO:HD2	1.99	0.44
1:A:296:TYR:CE2	1:A:299:ARG:HG2	2.52	0.44
1:B:97:PRO:HG2	1:B:100:ASP:HB2	1.99	0.44
1:A:75:MET:HE3	1:A:78:ARG:HD2	2.00	0.43
1:B:211:THR:HG22	2:B:1014:HOH:O	2.19	0.42
1:B:247:VAL:CG2	1:B:271:ILE:HD13	2.49	0.42
1:B:96:ALA:N	1:B:97:PRO:CD	2.83	0.41
1:A:326:HIS:CD2	1:A:383:ALA:HB2	2.56	0.41
1:A:572:GLY:O	1:A:573:PRO:C	2.59	0.41
1:A:96:ALA:N	1:A:97:PRO:CD	2.84	0.41

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	454/916 (50%)	437 (96%)	15 (3%)	2 (0%)	34	38
1	B	458/916 (50%)	440 (96%)	16 (4%)	2 (0%)	34	38
All	All	912/1832 (50%)	877 (96%)	31 (3%)	4 (0%)	34	38

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	314	PRO
1	B	314	PRO
1	B	712	SER
1	A	710	PRO

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	416/797 (52%)	388 (93%)	28 (7%)	16	17
1	B	419/797 (53%)	399 (95%)	20 (5%)	25	30
All	All	835/1594 (52%)	787 (94%)	48 (6%)	20	22

All (48) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	75	MET
1	A	101	TYR
1	A	142	ASP
1	A	149	LYS
1	A	150	GLU
1	A	167	ARG
1	A	186	GLU
1	A	235	ARG
1	A	273	ARG
1	A	278	TRP
1	A	289	LYS
1	A	299	ARG
1	A	301	THR
1	A	303	LEU
1	A	308	LYS
1	A	381	LYS
1	A	539	VAL
1	A	599	LYS
1	A	622	MET

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	626	THR
1	A	634	ARG
1	A	648	LYS
1	A	656	SER
1	A	676	GLN
1	A	677	GLU
1	A	678	PRO
1	A	682	ARG
1	A	709	THR
1	B	101	TYR
1	B	119	LEU
1	B	156	ARG
1	B	167	ARG
1	B	211	THR
1	B	273	ARG
1	B	278	TRP
1	B	301	THR
1	B	308	LYS
1	B	537	GLN
1	B	549	THR
1	B	577	GLN
1	B	608	ARG
1	B	613	LYS
1	B	648	LYS
1	B	653	MET
1	B	656	SER
1	B	675	ASP
1	B	677	GLU
1	B	709	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (9) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	103	HIS
1	A	193	HIS
1	A	326	HIS
1	A	577	GLN
1	A	630	ASN
1	A	676	GLN
1	B	103	HIS
1	B	163	GLN
1	B	193	HIS

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 carbohydrates 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, 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	466/916 (50%)	0.20	12 (2%)	56	65	57, 83, 124, 164	0
1	B	470/916 (51%)	0.21	19 (4%)	38	51	45, 68, 127, 160	0
All	All	936/1832 (51%)	0.21	31 (3%)	46	59	45, 78, 126, 164	0

All (31) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	673	ILE	6.1
1	B	669	ALA	5.2
1	B	665	SER	4.0
1	B	679	TRP	4.0
1	B	371	ASP	3.9
1	B	587	ALA	3.3
1	A	679	TRP	3.1
1	B	675	ASP	3.1
1	B	59	ARG	3.0
1	A	605	GLU	3.0
1	B	666	HIS	3.0
1	B	340	ARG	3.0
1	B	670	GLN	2.9
1	A	599	LYS	2.9
1	A	225	TYR	2.8
1	A	670	GLN	2.8
1	B	603	LYS	2.8
1	B	595	ASN	2.7
1	A	669	ALA	2.7
1	B	605	GLU	2.7
1	B	664	ARG	2.7
1	A	664	ARG	2.6
1	A	668	SER	2.6
1	B	663	CYS	2.5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	678	PRO	2.5
1	B	667	SER	2.4
1	B	668	SER	2.3
1	A	594	VAL	2.2
1	A	595	ASN	2.2
1	B	339	VAL	2.1
1	B	676	GLN	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 carbohydrates 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.