Full wwPDB X-ray Structure Validation Report (i)

May 13, 2020 – 09:24 am BST

PDB ID	:	5F8Q
Title	:	Blood group antigen binding adhesin BabA of Helicobacter pylori strain S831 in complex with Nanobody Nb-ER19
Authors	:	Moonens, K.; Gideonsson, P.; Subedi, S.; Romao, E.; Oscarson, S.; Muyldermans, S.; Boren, T.; Remaut, H.
Deposited on	:	2015-12-09
Resolution	:	2.59 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 (i) symbol.

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	:	2.11
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 2.59 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.
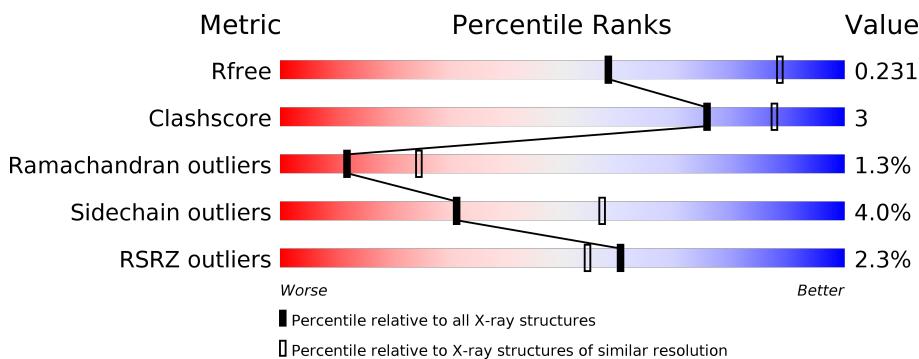
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

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	466	3% / 82% / 7% · 10%
1	B	466	3% / 80% / 8% · 10%
2	C	120	87% / 6% · 6%
2	D	120	92% / · 5%
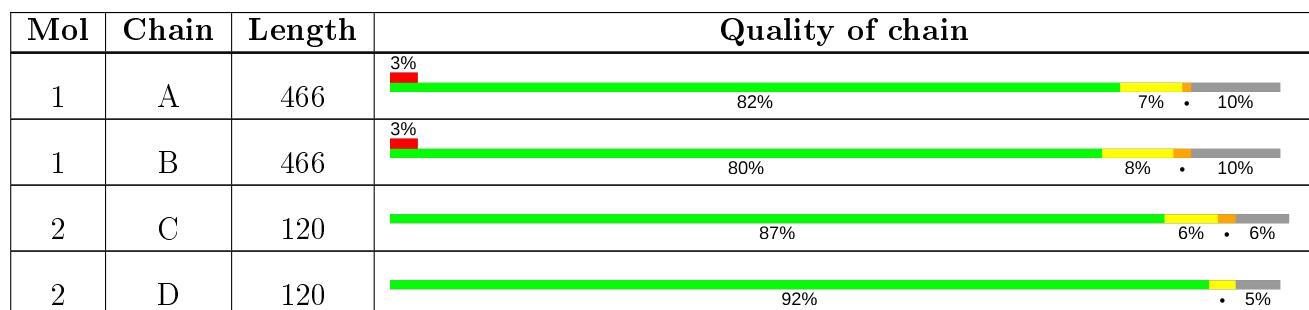
2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 8024 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 Adhesin binding fucosylated histo-blood group antigen,Adhesin,Adhesin binding fucosylated histo-blood group antigen.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	419	Total 3118	C 1925	N 537	O 643	S 13	0	0	0
1	B	419	Total 3118	C 1925	N 537	O 643	S 13	0	0	0

There are 60 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	3	ALA	-	expression tag	UNP O52269
A	4	SER	-	expression tag	UNP O52269
A	5	TRP	-	expression tag	UNP O52269
A	6	SER	-	expression tag	UNP O52269
A	7	HIS	-	expression tag	UNP O52269
A	8	PRO	-	expression tag	UNP O52269
A	9	GLN	-	expression tag	UNP O52269
A	10	PHE	-	expression tag	UNP O52269
A	11	GLU	-	expression tag	UNP O52269
A	12	LYS	-	expression tag	UNP O52269
A	13	SER	-	expression tag	UNP O52269
A	14	GLY	-	expression tag	UNP O52269
A	15	GLY	-	expression tag	UNP O52269
A	16	GLY	-	expression tag	UNP O52269
A	17	GLY	-	expression tag	UNP O52269
A	18	GLY	-	expression tag	UNP O52269
A	19	LEU	-	expression tag	UNP O52269
A	20	VAL	-	expression tag	UNP O52269
A	21	PRO	-	expression tag	UNP O52269
A	22	ARG	-	expression tag	UNP O52269
A	23	GLY	-	expression tag	UNP O52269
A	24	SER	-	expression tag	UNP O52269
A	461	GLY	-	expression tag	UNP O52269
A	462	SER	-	expression tag	UNP O52269

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	463	HIS	-	expression tag	UNP O52269
A	464	HIS	-	expression tag	UNP O52269
A	465	HIS	-	expression tag	UNP O52269
A	466	HIS	-	expression tag	UNP O52269
A	467	HIS	-	expression tag	UNP O52269
A	468	HIS	-	expression tag	UNP O52269
B	3	ALA	-	expression tag	UNP O52269
B	4	SER	-	expression tag	UNP O52269
B	5	TRP	-	expression tag	UNP O52269
B	6	SER	-	expression tag	UNP O52269
B	7	HIS	-	expression tag	UNP O52269
B	8	PRO	-	expression tag	UNP O52269
B	9	GLN	-	expression tag	UNP O52269
B	10	PHE	-	expression tag	UNP O52269
B	11	GLU	-	expression tag	UNP O52269
B	12	LYS	-	expression tag	UNP O52269
B	13	SER	-	expression tag	UNP O52269
B	14	GLY	-	expression tag	UNP O52269
B	15	GLY	-	expression tag	UNP O52269
B	16	GLY	-	expression tag	UNP O52269
B	17	GLY	-	expression tag	UNP O52269
B	18	GLY	-	expression tag	UNP O52269
B	19	LEU	-	expression tag	UNP O52269
B	20	VAL	-	expression tag	UNP O52269
B	21	PRO	-	expression tag	UNP O52269
B	22	ARG	-	expression tag	UNP O52269
B	23	GLY	-	expression tag	UNP O52269
B	24	SER	-	expression tag	UNP O52269
B	461	GLY	-	expression tag	UNP O52269
B	462	SER	-	expression tag	UNP O52269
B	463	HIS	-	expression tag	UNP O52269
B	464	HIS	-	expression tag	UNP O52269
B	465	HIS	-	expression tag	UNP O52269
B	466	HIS	-	expression tag	UNP O52269
B	467	HIS	-	expression tag	UNP O52269
B	468	HIS	-	expression tag	UNP O52269

- Molecule 2 is a protein called Nanobody Nb-ER19.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	C	113	863	540	155	163	5	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	D	114	873	546	158	164	5	0	0	0

- Molecule 3 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	21	Total 21	O 21	0	0
3	C	4	Total 4	O 4	0	0
3	B	17	Total 17	O 17	0	0
3	D	10	Total 10	O 10	0	0

3 Residue-property plots ⓘ

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: Adhesin binding fucosylated histo-blood group antigen,Adhesin,Adhesin binding fucosylated histo-blood group antigen
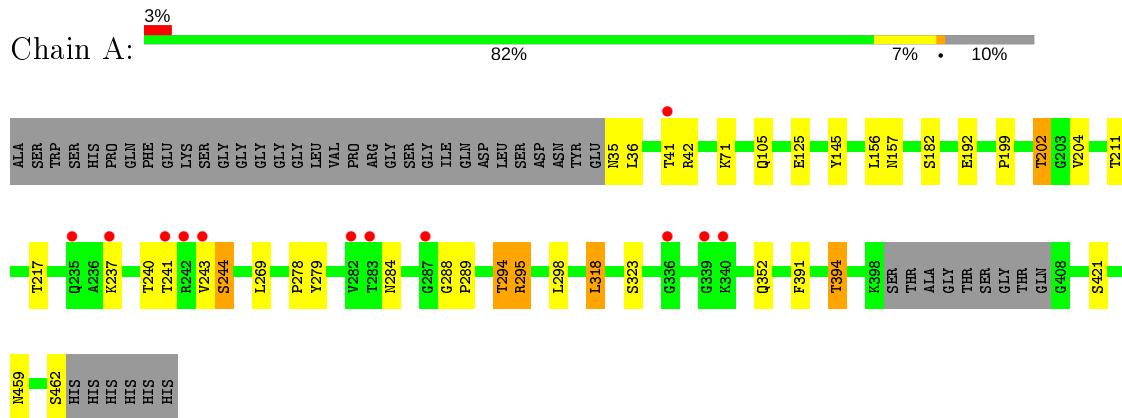
• Molecule 1: Adhesin binding fucosylated histo-blood group antigen,Adhesin,Adhesin binding fucosylated histo-blood group antigen
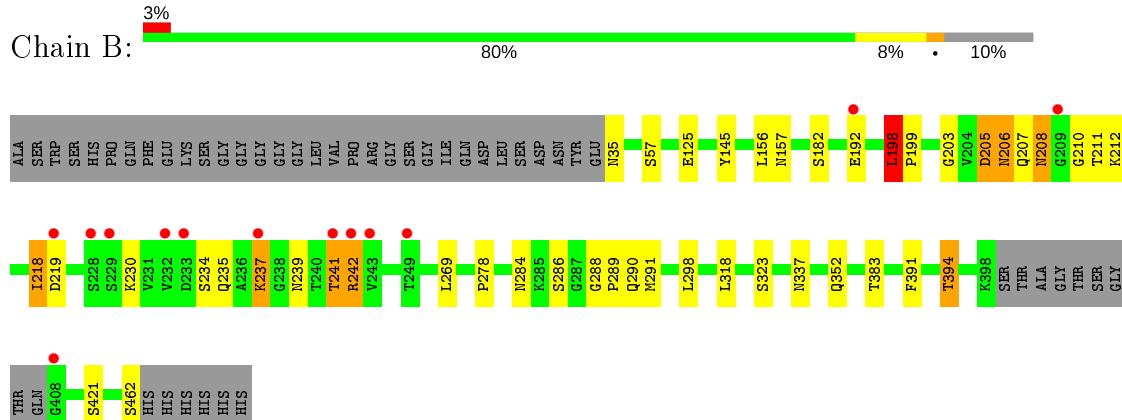
• Molecule 2: Nanobody Nb-ER19
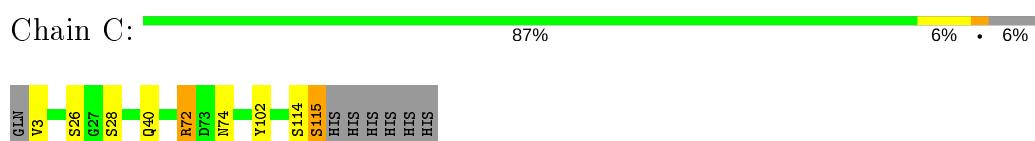
• Molecule 2: Nanobody Nb-ER19

Chain D:

92%   •   5%

GLN V3 S28 Q40 R72 H116 HIS HIS HIS HIS HIS
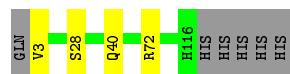
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	51.03Å 132.14Å 123.48Å 90.00° 94.90° 90.00°	Depositor
Resolution (Å)	48.50 – 2.59 47.46 – 2.59	Depositor EDS
% Data completeness (in resolution range)	97.5 (48.50-2.59) 97.7 (47.46-2.59)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.18 (at 2.58Å)	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
R, R_{free}	0.195 , 0.230 0.198 , 0.231	Depositor DCC
R_{free} test set	2487 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	65.9	Xtriage
Anisotropy	0.153	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 33.8	EDS
L-test for twinning[2]	$<|L|> = 0.46$, $<L^2> = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	8024	wwPDB-VP
Average B, all atoms (Å^2)	68.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.71% 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 (i)

5.1 Standard geometry (i)

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	0.71	0/3165	0.76	0/4305
1	B	0.64	0/3165	0.76	1/4305 (0.0%)
2	C	0.66	0/882	0.83	2/1195 (0.2%)
2	D	0.71	0/893	0.82	1/1210 (0.1%)
All	All	0.68	0/8105	0.78	4/11015 (0.0%)

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	B	0	1
2	D	0	1
All	All	0	2

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	198	LEU	CA-CB-CG	9.12	136.27	115.30
2	C	72	ARG	NE-CZ-NH1	7.26	123.93	120.30
2	C	72	ARG	NE-CZ-NH2	-5.10	117.75	120.30
2	D	72	ARG	NE-CZ-NH1	5.09	122.85	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	286	SER	Peptide
2	D	3	VAL	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	3118	0	3049	20	0
1	B	3118	0	3049	20	0
2	C	863	0	845	9	0
2	D	873	0	852	0	0
3	A	21	0	0	1	0
3	B	17	0	0	1	0
3	C	4	0	0	0	0
3	D	10	0	0	0	0
All	All	8024	0	7795	49	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 3.

All (49) 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:284:ASN:HB2	3:A:504:HOH:O	1.56	1.04
2:C:3:VAL:HG13	2:C:102:TYR:CZ	2.14	0.83
2:C:3:VAL:HG11	2:C:102:TYR:OH	1.81	0.80
2:C:3:VAL:N	2:C:26:SER:HG	1.80	0.80
2:C:3:VAL:HG13	2:C:102:TYR:CE2	2.19	0.77
2:C:3:VAL:CG1	2:C:102:TYR:OH	2.42	0.67
1:B:210:GLY:HA3	1:B:211:THR:HG22	1.76	0.66
2:C:3:VAL:CG1	2:C:102:TYR:CZ	2.78	0.65
1:B:205:ASP:O	1:B:206:ASN:ND2	2.32	0.62
1:A:279:TYR:OH	1:A:295:ARG:CZ	2.48	0.61
2:C:72:ARG:HD3	2:C:74:ASN:OD1	2.01	0.60
1:A:35:ASN:OD1	1:A:35:ASN:C	2.40	0.57
1:B:241:THR:O	1:B:242:ARG:CB	2.53	0.57
1:B:35:ASN:C	1:B:35:ASN:OD1	2.43	0.57
2:C:114:SER:O	2:C:115:SER:HB2	2.03	0.57
1:A:279:TYR:OH	1:A:295:ARG:NE	2.38	0.56
1:A:202:THR:CG2	1:A:204:VAL:HB	2.40	0.52
1:A:156:LEU:HG	1:A:269:LEU:HD12	1.92	0.51
1:B:284:ASN:HD21	1:B:291:MET:HB3	1.75	0.51

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:294:THR:O	1:A:295:ARG:O	2.30	0.50
1:B:156:LEU:HG	1:B:269:LEU:HD12	1.94	0.49
1:B:198:LEU:HB2	1:B:199:PRO:HD2	1.94	0.48
2:C:3:VAL:O	2:C:3:VAL:HG12	2.13	0.48
1:B:218:ILE:O	1:B:218:ILE:HG13	2.14	0.48
1:B:239:ASN:HB3	3:B:501:HOH:O	2.12	0.47
1:A:105:GLN:CD	1:A:279:TYR:HD2	2.18	0.47
1:A:394:THR:HG22	1:A:421:SER:HB2	1.95	0.47
1:B:205:ASP:C	1:B:206:ASN:ND2	2.68	0.47
1:B:284:ASN:OD1	1:B:290:GLN:HA	2.15	0.47
1:B:394:THR:HG22	1:B:421:SER:HB2	1.98	0.47
1:B:278:PRO:HG2	1:B:298:LEU:HD22	1.97	0.46
1:B:208:ASN:HA	1:B:230:LYS:HA	1.98	0.46
1:A:243:VAL:HG12	1:A:244:SER:N	2.32	0.45
1:B:203:GLY:O	1:B:212:LYS:HD3	2.16	0.45
1:B:234:SER:O	1:B:235:GLN:HB2	2.16	0.45
1:B:125:GLU:OE2	1:B:157:ASN:OD1	2.35	0.45
1:A:278:PRO:HG2	1:A:298:LEU:HD22	1.98	0.45
1:B:241:THR:O	1:B:242:ARG:HB2	2.17	0.44
1:A:279:TYR:OH	1:A:295:ARG:CG	2.66	0.44
1:A:240:THR:O	1:A:241:THR:C	2.56	0.43
1:B:199:PRO:HD3	1:B:207:GLN:HG3	2.01	0.43
1:A:217:THR:HG22	1:A:217:THR:O	2.19	0.42
1:A:391:PHE:O	1:A:394:THR:HB	2.19	0.42
1:A:294:THR:O	1:A:295:ARG:C	2.57	0.42
1:A:35:ASN:OD1	1:A:36:LEU:N	2.53	0.42
1:A:125:GLU:OE2	1:A:157:ASN:OD1	2.37	0.42
1:A:318:LEU:HD23	1:A:318:LEU:C	2.40	0.42
1:A:105:GLN:CD	1:A:279:TYR:CD2	2.95	0.41
1:B:391:PHE:O	1:B:394:THR:HB	2.21	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	415/466 (89%)	391 (94%)	18 (4%)	6 (1%)	11	22
1	B	415/466 (89%)	393 (95%)	14 (3%)	8 (2%)	8	15
2	C	111/120 (92%)	108 (97%)	3 (3%)	0	100	100
2	D	112/120 (93%)	110 (98%)	2 (2%)	0	100	100
All	All	1053/1172 (90%)	1002 (95%)	37 (4%)	14 (1%)	12	24

All (14) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	237	LYS
1	A	295	ARG
1	A	288	GLY
1	B	219	ASP
1	B	242	ARG
1	A	294	THR
1	B	208	ASN
1	B	237	LYS
1	B	241	THR
1	B	288	GLY
1	A	41	THR
1	A	289	PRO
1	B	206	ASN
1	B	289	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	346/382 (91%)	331 (96%)	15 (4%)	29	54
1	B	346/382 (91%)	331 (96%)	15 (4%)	29	54
2	C	92/99 (93%)	89 (97%)	3 (3%)	38	64

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	D	93/99 (94%)	91 (98%)	2 (2%)	52	76
All	All	877/962 (91%)	842 (96%)	35 (4%)	31	57

All (35) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	42	ARG
1	A	71	LYS
1	A	145	TYR
1	A	182	SER
1	A	192	GLU
1	A	199	PRO
1	A	202	THR
1	A	211	THR
1	A	244	SER
1	A	318	LEU
1	A	323	SER
1	A	352	GLN
1	A	394	THR
1	A	459	ASN
1	A	462	SER
2	C	28	SER
2	C	40	GLN
2	C	115	SER
1	B	57	SER
1	B	145	TYR
1	B	182	SER
1	B	192	GLU
1	B	198	LEU
1	B	205	ASP
1	B	218	ILE
1	B	237	LYS
1	B	318	LEU
1	B	323	SER
1	B	337	ASN
1	B	352	GLN
1	B	383	THR
1	B	394	THR
1	B	462	SER
2	D	28	SER
2	D	40	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	352	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 carbohydrates 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 (i)

6.1 Protein, DNA and RNA chains (i)

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	419/466 (89%)	0.13	12 (2%)	51	45	40, 61, 109, 135	0
1	B	419/466 (89%)	0.24	13 (3%)	49	42	45, 68, 120, 153	0
2	C	113/120 (94%)	-0.01	0	100	100	49, 63, 84, 102	0
2	D	114/120 (95%)	0.05	0	100	100	45, 55, 74, 99	0
All	All	1065/1172 (90%)	0.15	25 (2%)	60	54	40, 63, 110, 153	0

All (25) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	219	ASP	3.7
1	B	232	VAL	3.5
1	B	241	THR	3.2
1	B	243	VAL	3.2
1	B	209	GLY	3.1
1	B	408	GLY	3.1
1	A	237	LYS	3.1
1	A	336	GLY	3.0
1	B	192	GLU	2.8
1	A	339	GLY	2.5
1	A	241	THR	2.5
1	A	340	LYS	2.5
1	A	282	VAL	2.4
1	A	283	THR	2.4
1	B	229	SER	2.4
1	A	242	ARG	2.3
1	A	235	GLN	2.3
1	B	228	SER	2.3
1	B	242	ARG	2.2
1	B	249	THR	2.2
1	A	287	GLY	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	41	THR	2.0
1	B	233	ASP	2.0
1	A	243	VAL	2.0
1	B	237	LYS	2.0

6.2 Non-standard residues in protein, DNA, RNA chains (i)

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

6.3 Carbohydrates (i)

There are no carbohydrates in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.