WORLDWIDE PDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 12:10 AM GMT

PDB ID : 1ZXO

Title : X-ray Crystal Structure of Protein Q8A1P1 from Bacteroides thetaiotaomi-
cron. Northeast Structural Genomics Consortium Target BtR25.

Authors : Kuzin, A.P.; Yong, W.; Forouhar, F.; Vorobiev, S.; Xiao, R.; Ma, C.; Ac-
ton, T.; Montelione, G.T.; Hunt, J.; Tong, L.; Northeast Structural Genomics
Consortium (NESG)

Deposited on : 2005-06-08

Resolution : 3.20 Å(reported)

The following versions of software and data (see references (1)) were used in the production of this report:

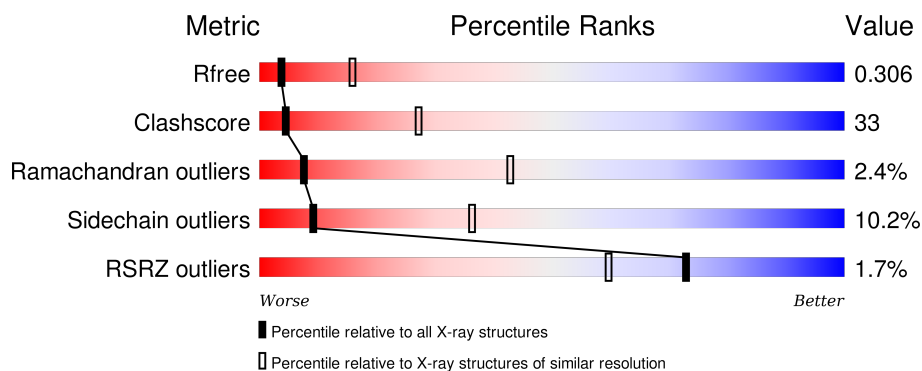|		
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 3.20 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		0.306
Clashscore		33
Ramachandran outliers		2.4%
Sidechain outliers		10.2%
RSRZ outliers		1.7%

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}	91344	1124 (3.24-3.16)
Clashscore	102246	1024 (3.22-3.18)
Ramachandran outliers	100387	1004 (3.22-3.18)
Sidechain outliers	100360	1003 (3.22-3.18)
RSRZ outliers	91569	1129 (3.24-3.16)

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. 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	291	7% — 40% / 46% / 5% / 9%
1	B	291	44% / 45% / 7% / •
1	C	291	48% / 42% / 6% / •
1	D	291	46% / 45% / 5% / •
1	E	291	44% / 44% / 8% / • •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	291	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 13033 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 conserved hypothetical protein Q8A1P1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	265	Total 2050	C 1319	N 346	O 375	S 6	Se 4	0	0	0
1	F	274	Total 2124	C 1368	N 358	O 387	S 6	Se 5	0	0	0
1	B	280	Total 2166	C 1395	N 365	O 395	S 6	Se 5	0	0	0
1	D	280	Total 2166	C 1395	N 365	O 395	S 6	Se 5	0	0	0
1	C	280	Total 2166	C 1395	N 365	O 395	S 6	Se 5	0	0	0
1	E	280	Total 2166	C 1395	N 365	O 395	S 6	Se 5	0	0	0

There are 78 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
A	96	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
A	215	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
A	227	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
A	269	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
A	284	LEU	-	EXPRESSION TAG	UNP Q8A1P1
A	285	GLU	-	EXPRESSION TAG	UNP Q8A1P1
A	286	HIS	-	EXPRESSION TAG	UNP Q8A1P1
A	287	HIS	-	EXPRESSION TAG	UNP Q8A1P1
A	288	HIS	-	EXPRESSION TAG	UNP Q8A1P1
A	289	HIS	-	EXPRESSION TAG	UNP Q8A1P1
A	290	HIS	-	EXPRESSION TAG	UNP Q8A1P1
A	291	HIS	-	EXPRESSION TAG	UNP Q8A1P1
B	1	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
B	96	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
B	215	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
B	227	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	269	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
B	284	LEU	-	EXPRESSION TAG	UNP Q8A1P1
B	285	GLU	-	EXPRESSION TAG	UNP Q8A1P1
B	286	HIS	-	EXPRESSION TAG	UNP Q8A1P1
B	287	HIS	-	EXPRESSION TAG	UNP Q8A1P1
B	288	HIS	-	EXPRESSION TAG	UNP Q8A1P1
B	289	HIS	-	EXPRESSION TAG	UNP Q8A1P1
B	290	HIS	-	EXPRESSION TAG	UNP Q8A1P1
B	291	HIS	-	EXPRESSION TAG	UNP Q8A1P1
C	1	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
C	96	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
C	215	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
C	227	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
C	269	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
C	284	LEU	-	EXPRESSION TAG	UNP Q8A1P1
C	285	GLU	-	EXPRESSION TAG	UNP Q8A1P1
C	286	HIS	-	EXPRESSION TAG	UNP Q8A1P1
C	287	HIS	-	EXPRESSION TAG	UNP Q8A1P1
C	288	HIS	-	EXPRESSION TAG	UNP Q8A1P1
C	289	HIS	-	EXPRESSION TAG	UNP Q8A1P1
C	290	HIS	-	EXPRESSION TAG	UNP Q8A1P1
C	291	HIS	-	EXPRESSION TAG	UNP Q8A1P1
D	1	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
D	96	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
D	215	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
D	227	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
D	269	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
D	284	LEU	-	EXPRESSION TAG	UNP Q8A1P1
D	285	GLU	-	EXPRESSION TAG	UNP Q8A1P1
D	286	HIS	-	EXPRESSION TAG	UNP Q8A1P1
D	287	HIS	-	EXPRESSION TAG	UNP Q8A1P1
D	288	HIS	-	EXPRESSION TAG	UNP Q8A1P1
D	289	HIS	-	EXPRESSION TAG	UNP Q8A1P1
D	290	HIS	-	EXPRESSION TAG	UNP Q8A1P1
D	291	HIS	-	EXPRESSION TAG	UNP Q8A1P1
E	1	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
E	96	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
E	215	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
E	227	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
E	269	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
E	284	LEU	-	EXPRESSION TAG	UNP Q8A1P1
E	285	GLU	-	EXPRESSION TAG	UNP Q8A1P1

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	286	HIS	-	EXPRESSION TAG	UNP Q8A1P1
E	287	HIS	-	EXPRESSION TAG	UNP Q8A1P1
E	288	HIS	-	EXPRESSION TAG	UNP Q8A1P1
E	289	HIS	-	EXPRESSION TAG	UNP Q8A1P1
E	290	HIS	-	EXPRESSION TAG	UNP Q8A1P1
E	291	HIS	-	EXPRESSION TAG	UNP Q8A1P1
F	1	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
F	96	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
F	215	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
F	227	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
F	269	MSE	MET	MODIFIED RESIDUE	UNP Q8A1P1
F	284	LEU	-	EXPRESSION TAG	UNP Q8A1P1
F	285	GLU	-	EXPRESSION TAG	UNP Q8A1P1
F	286	HIS	-	EXPRESSION TAG	UNP Q8A1P1
F	287	HIS	-	EXPRESSION TAG	UNP Q8A1P1
F	288	HIS	-	EXPRESSION TAG	UNP Q8A1P1
F	289	HIS	-	EXPRESSION TAG	UNP Q8A1P1
F	290	HIS	-	EXPRESSION TAG	UNP Q8A1P1
F	291	HIS	-	EXPRESSION TAG	UNP Q8A1P1

- Molecule 2 is water.

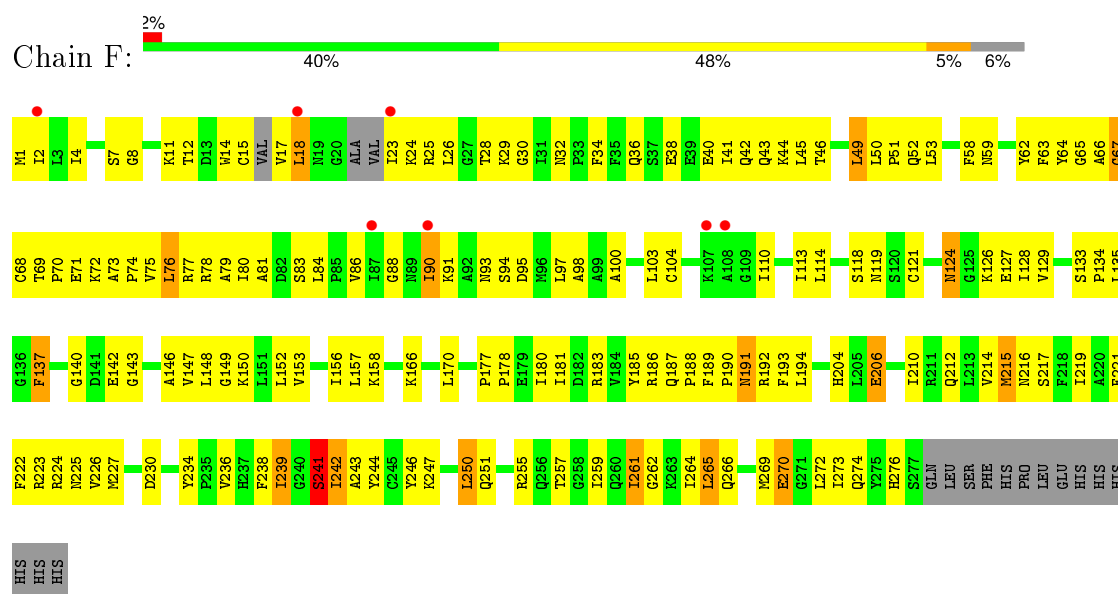| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	16	Total 16	O 16	0	0
2	B	34	Total 34	O 34	0	0
2	C	40	Total 40	O 40	0	0
2	D	32	Total 32	O 32	0	0
2	E	36	Total 36	O 36	0	0
2	F	37	Total 37	O 37	0	0

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 errors 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.
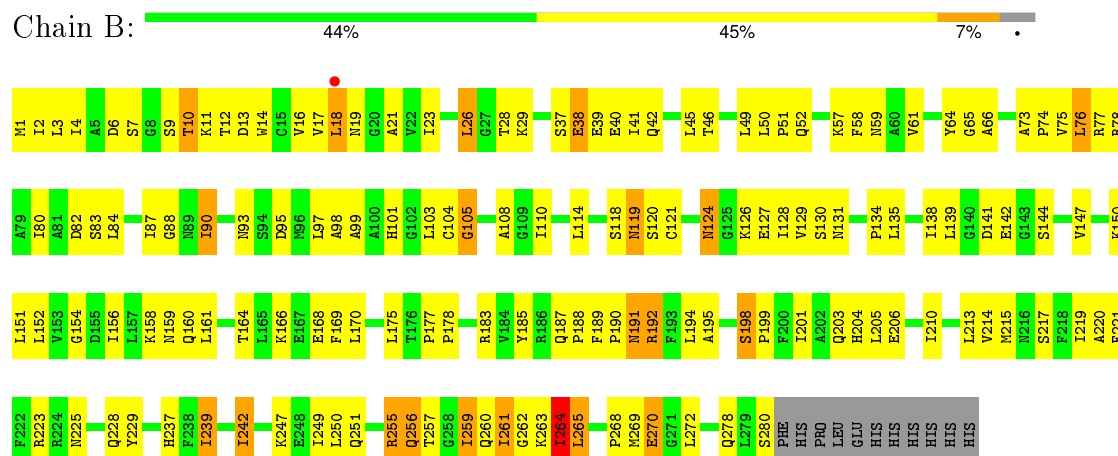
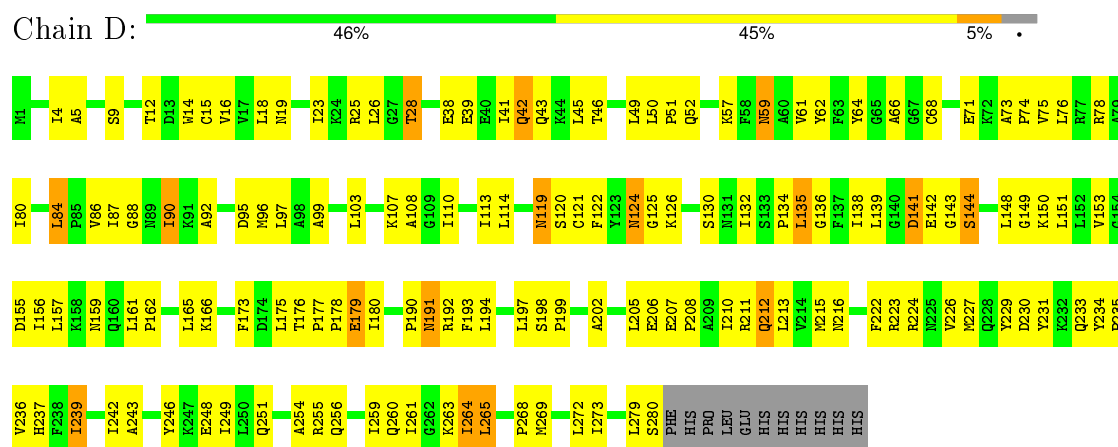• Molecule 1: conserved hypothetical protein Q8A1P1

• Molecule 1: conserved hypothetical protein Q8A1P1

• Molecule 1: conserved hypothetical protein Q8A1P1

Chain B:

• Molecule 1: conserved hypothetical protein Q8A1P1

Chain D:

• Molecule 1: conserved hypothetical protein Q8A1P1

Chain C:

• Molecule 1: conserved hypothetical protein Q8A1P1

Chain E:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, α, β, γ	158.15Å 158.15Å 275.53Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	25.56 – 3.20 29.89 – 3.00	Depositor EDS
% Data completeness (in resolution range)	92.2 (25.56-3.20) 98.2 (29.89-3.00)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	3.00 (at 3.00Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.241 , 0.297 0.249 , 0.306	Depositor DCC
R_{free} test set	2042 reflections (4.87%)	DCC
Wilson B-factor (Å^2)	58.8	Xtriage
Anisotropy	0.214	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 11.0	EDS
Estimated twinning fraction	0.447 for h,-h-k,-l	Xtriage
L-test for twinning[2]	$<|L|> = 0.49, <L^2> = 0.32$	Xtriage
Outliers	0 of 102631 reflections	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	13033	wwPDB-VP
Average B, all atoms (Å^2)	60.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.86% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.34	0/2088	0.55	0/2816
1	B	0.48	0/2208	0.65	0/2981
1	C	0.48	0/2208	0.68	0/2981
1	D	0.49	0/2208	0.66	0/2981
1	E	0.50	0/2208	0.68	1/2981 (0.0%)
1	F	0.39	0/2164	0.57	0/2917
All	All	0.45	0/13084	0.64	1/17657 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	E	25	ARG	NE-CZ-NH1	-5.13	117.73	120.30

There are no chirality outliers.

There are no planarity outliers.

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	2050	0	2076	136	0
1	B	2166	0	2203	152	0
1	C	2166	0	2203	135	0
1	D	2166	0	2203	124	0
1	E	2166	0	2203	164	0
1	F	2124	0	2154	155	0

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	16	0	0	0	0
2	B	34	0	0	3	0
2	C	40	0	0	4	0
2	D	32	0	0	2	0
2	E	36	0	0	5	0
2	F	37	0	0	11	0
All	All	13033	0	13042	842	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 33.

The worst 5 of 842 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:E:25:ARG:NH1	2:E:295:HOH:O	1.73	1.15
1:E:25:ARG:CZ	2:E:295:HOH:O	1.90	1.11
1:B:1:MSE:HB3	1:B:18:LEU:HD22	1.29	1.08
1:A:135:LEU:H	1:A:135:LEU:HD22	1.24	1.02
1:C:54:PRO:O	2:C:302:HOH:O	1.79	0.98

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	261/291 (90%)	206 (79%)	48 (18%)	7 (3%)	6	39
1	B	278/291 (96%)	230 (83%)	39 (14%)	9 (3%)	5	33
1	C	278/291 (96%)	243 (87%)	35 (13%)	0	100	100
1	D	278/291 (96%)	238 (86%)	35 (13%)	5 (2%)	11	51
1	E	278/291 (96%)	232 (84%)	40 (14%)	6 (2%)	8	45

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	F	268/291 (92%)	215 (80%)	41 (15%)	12 (4%)	3	24
All	All	1641/1746 (94%)	1364 (83%)	238 (14%)	39 (2%)	7	43

5 of 39 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	10	THR
1	F	241	SER
1	E	264	ILE
1	A	9	SER
1	A	40	GLU

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	219/239 (92%)	201 (92%)	18 (8%)	14	50
1	B	233/239 (98%)	207 (89%)	26 (11%)	7	32
1	C	233/239 (98%)	203 (87%)	30 (13%)	5	24
1	D	233/239 (98%)	212 (91%)	21 (9%)	12	43
1	E	233/239 (98%)	202 (87%)	31 (13%)	5	23
1	F	228/239 (95%)	214 (94%)	14 (6%)	23	64
All	All	1379/1434 (96%)	1239 (90%)	140 (10%)	9	36

5 of 140 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	84	LEU
1	D	265	LEU
1	E	198	SER
1	D	119	ASN
1	D	191	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 94 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	260	GLN
1	D	212	GLN
1	E	204	HIS
1	B	274	GLN
1	D	124	ASN

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(Å^2)	Q<0.9
1	A	261/291 (89%)	0.49	19 (7%)	18	10	62, 112, 140, 157	0
1	B	275/291 (94%)	-0.27	1 (0%)	93	90	12, 37, 88, 142	0
1	C	275/291 (94%)	-0.30	0	100	100	4, 34, 65, 114	0
1	D	275/291 (94%)	-0.33	0	100	100	9, 33, 67, 119	0
1	E	275/291 (94%)	-0.30	0	100	100	9, 35, 80, 116	0
1	F	269/291 (92%)	0.25	7 (2%)	59	45	54, 96, 134, 145	0
All	All	1630/1746 (93%)	-0.08	27 (1%)	73	60	4, 49, 129, 157	0

The worst 5 of 27 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	16	VAL	4.4
1	A	280	SER	4.4
1	A	235	PRO	3.6
1	B	18	LEU	3.6
1	F	2	ILE	3.4

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.