wwPDB X-ray Structure Validation Summary Report (i)

Oct 12, 2022 – 12:11 PM EDT

PDB ID	:	7T31
Title	:	X-ray Structure of Clostridiodies difficile PilW
Authors	:	Ronish, L.A.; Piepenbrink, K.H.
Deposited on	:	2021-12-06
Resolution	:	2.30 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:

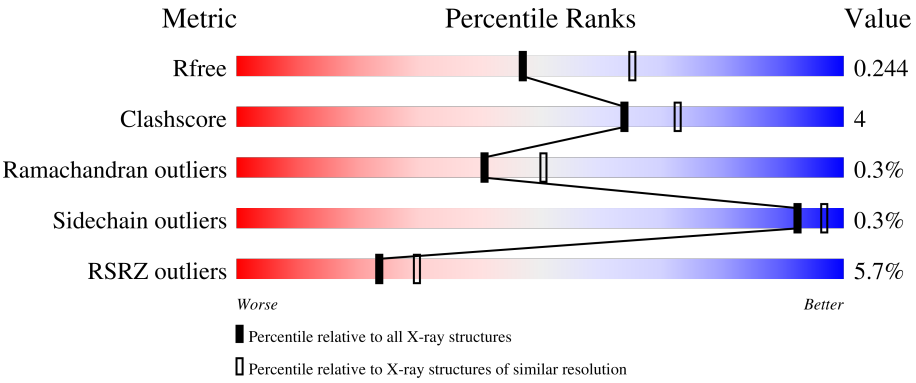
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.31.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.2

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 2.30 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	5042 (2.30-2.30)
Clashscore	141614	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)
RSRZ outliers	127900	4938 (2.30-2.30)

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	512	6% 84% 12% ·
1	B	512	6% 88% 8% ·
1	C	512	4% 88% 9% · ·
1	D	512	6% 87% 7% 6%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 16068 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 Maltose/maltodextrin-binding periplasmic protein,Putative pilin protein chimera.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	490	Total	C	N	O	S	0	3	0
			3731	2397	604	720	10			
1	B	497	Total	C	N	O	S	0	4	0
			3810	2449	619	732	10			
1	C	497	Total	C	N	O	S	0	3	0
			3824	2456	619	740	9			
1	D	482	Total	C	N	O	S	0	5	0
			3708	2389	597	711	11			

There are 112 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	MET	-	initiating methionine	UNP P0AEX9
A	82	ALA	ASP	conflict	UNP P0AEX9
A	83	ALA	LYS	conflict	UNP P0AEX9
A	172	ALA	GLU	conflict	UNP P0AEX9
A	173	ALA	ASN	conflict	UNP P0AEX9
A	239	ALA	LYS	conflict	UNP P0AEX9
A	359	ALA	GLU	conflict	UNP P0AEX9
A	362	ALA	LYS	conflict	UNP P0AEX9
A	363	ALA	ASP	conflict	UNP P0AEX9
A	367	ASN	ARG	conflict	UNP P0AEX9
A	368	ALA	-	linker	UNP P0AEX9
A	369	ALA	-	linker	UNP P0AEX9
A	370	ALA	-	linker	UNP P0AEX9
A	1083	ALA	LEU	conflict	UNP Q185H8
A	1130	ASN	SER	conflict	UNP Q185H8
A	1134	ASN	SER	conflict	UNP Q185H8
A	1135	LEU	SER	conflict	UNP Q185H8
A	1139	THR	LYS	conflict	UNP Q185H8
A	1146	ASN	ASP	conflict	UNP Q185H8
A	1154	ASN	ASP	conflict	UNP Q185H8

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	1159	LEU	-	expression tag	UNP Q185H8
A	1160	GLU	-	expression tag	UNP Q185H8
A	1161	HIS	-	expression tag	UNP Q185H8
A	1162	HIS	-	expression tag	UNP Q185H8
A	1163	HIS	-	expression tag	UNP Q185H8
A	1164	HIS	-	expression tag	UNP Q185H8
A	1165	HIS	-	expression tag	UNP Q185H8
A	1166	HIS	-	expression tag	UNP Q185H8
B	0	MET	-	initiating methionine	UNP P0AEX9
B	82	ALA	ASP	conflict	UNP P0AEX9
B	83	ALA	LYS	conflict	UNP P0AEX9
B	172	ALA	GLU	conflict	UNP P0AEX9
B	173	ALA	ASN	conflict	UNP P0AEX9
B	239	ALA	LYS	conflict	UNP P0AEX9
B	359	ALA	GLU	conflict	UNP P0AEX9
B	362	ALA	LYS	conflict	UNP P0AEX9
B	363	ALA	ASP	conflict	UNP P0AEX9
B	367	ASN	ARG	conflict	UNP P0AEX9
B	368	ALA	-	linker	UNP P0AEX9
B	369	ALA	-	linker	UNP P0AEX9
B	370	ALA	-	linker	UNP P0AEX9
B	1083	ALA	LEU	conflict	UNP Q185H8
B	1130	ASN	SER	conflict	UNP Q185H8
B	1134	ASN	SER	conflict	UNP Q185H8
B	1135	LEU	SER	conflict	UNP Q185H8
B	1139	THR	LYS	conflict	UNP Q185H8
B	1146	ASN	ASP	conflict	UNP Q185H8
B	1154	ASN	ASP	conflict	UNP Q185H8
B	1159	LEU	-	expression tag	UNP Q185H8
B	1160	GLU	-	expression tag	UNP Q185H8
B	1161	HIS	-	expression tag	UNP Q185H8
B	1162	HIS	-	expression tag	UNP Q185H8
B	1163	HIS	-	expression tag	UNP Q185H8
B	1164	HIS	-	expression tag	UNP Q185H8
B	1165	HIS	-	expression tag	UNP Q185H8
B	1166	HIS	-	expression tag	UNP Q185H8
C	0	MET	-	initiating methionine	UNP P0AEX9
C	82	ALA	ASP	conflict	UNP P0AEX9
C	83	ALA	LYS	conflict	UNP P0AEX9
C	172	ALA	GLU	conflict	UNP P0AEX9
C	173	ALA	ASN	conflict	UNP P0AEX9
C	239	ALA	LYS	conflict	UNP P0AEX9

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	359	ALA	GLU	conflict	UNP P0AEX9
C	362	ALA	LYS	conflict	UNP P0AEX9
C	363	ALA	ASP	conflict	UNP P0AEX9
C	367	ASN	ARG	conflict	UNP P0AEX9
C	368	ALA	-	linker	UNP P0AEX9
C	369	ALA	-	linker	UNP P0AEX9
C	370	ALA	-	linker	UNP P0AEX9
C	1083	ALA	LEU	conflict	UNP Q185H8
C	1130	ASN	SER	conflict	UNP Q185H8
C	1134	ASN	SER	conflict	UNP Q185H8
C	1135	LEU	SER	conflict	UNP Q185H8
C	1139	THR	LYS	conflict	UNP Q185H8
C	1146	ASN	ASP	conflict	UNP Q185H8
C	1154	ASN	ASP	conflict	UNP Q185H8
C	1159	LEU	-	expression tag	UNP Q185H8
C	1160	GLU	-	expression tag	UNP Q185H8
C	1161	HIS	-	expression tag	UNP Q185H8
C	1162	HIS	-	expression tag	UNP Q185H8
C	1163	HIS	-	expression tag	UNP Q185H8
C	1164	HIS	-	expression tag	UNP Q185H8
C	1165	HIS	-	expression tag	UNP Q185H8
C	1166	HIS	-	expression tag	UNP Q185H8
D	0	MET	-	initiating methionine	UNP P0AEX9
D	82	ALA	ASP	conflict	UNP P0AEX9
D	83	ALA	LYS	conflict	UNP P0AEX9
D	172	ALA	GLU	conflict	UNP P0AEX9
D	173	ALA	ASN	conflict	UNP P0AEX9
D	239	ALA	LYS	conflict	UNP P0AEX9
D	359	ALA	GLU	conflict	UNP P0AEX9
D	362	ALA	LYS	conflict	UNP P0AEX9
D	363	ALA	ASP	conflict	UNP P0AEX9
D	367	ASN	ARG	conflict	UNP P0AEX9
D	368	ALA	-	linker	UNP P0AEX9
D	369	ALA	-	linker	UNP P0AEX9
D	370	ALA	-	linker	UNP P0AEX9
D	1083	ALA	LEU	conflict	UNP Q185H8
D	1130	ASN	SER	conflict	UNP Q185H8
D	1134	ASN	SER	conflict	UNP Q185H8
D	1135	LEU	SER	conflict	UNP Q185H8
D	1139	THR	LYS	conflict	UNP Q185H8
D	1146	ASN	ASP	conflict	UNP Q185H8
D	1154	ASN	ASP	conflict	UNP Q185H8

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	1159	LEU	-	expression tag	UNP Q185H8
D	1160	GLU	-	expression tag	UNP Q185H8
D	1161	HIS	-	expression tag	UNP Q185H8
D	1162	HIS	-	expression tag	UNP Q185H8
D	1163	HIS	-	expression tag	UNP Q185H8
D	1164	HIS	-	expression tag	UNP Q185H8
D	1165	HIS	-	expression tag	UNP Q185H8
D	1166	HIS	-	expression tag	UNP Q185H8

- Molecule 2 is water.

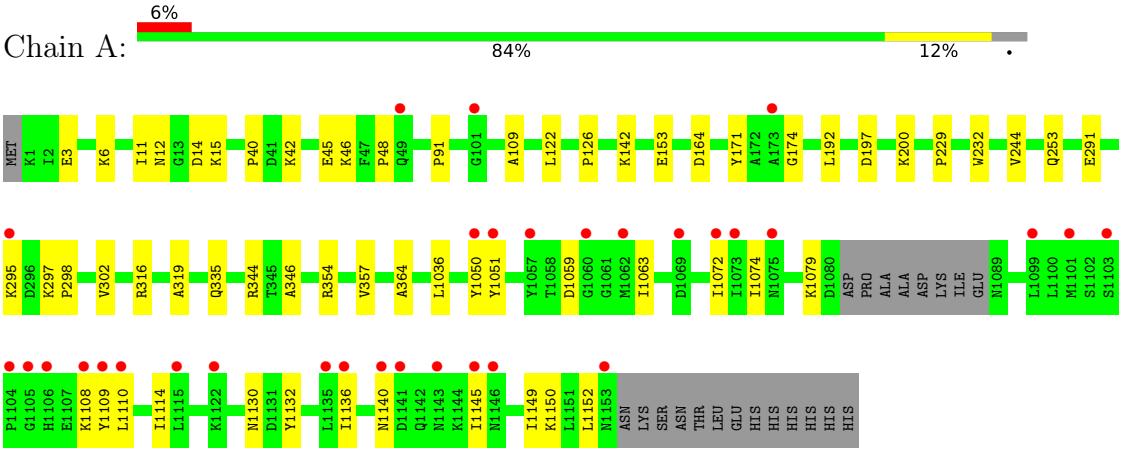
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	160	Total O 160 160	0	0
2	B	270	Total O 270 270	0	0
2	C	292	Total O 292 292	0	0
2	D	273	Total O 273 273	0	0

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.

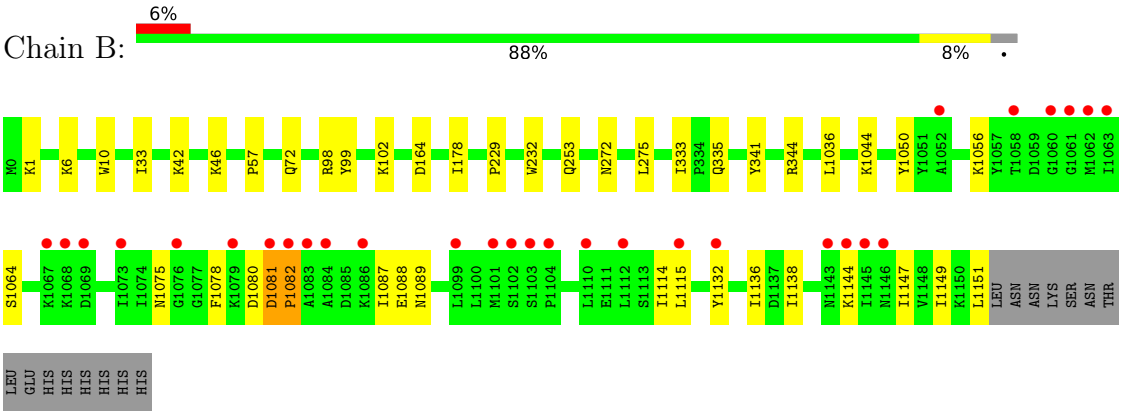
• Molecule 1: Maltose/maltodextrin-binding periplasmic protein,Putative pilin protein chimera

Chain A:



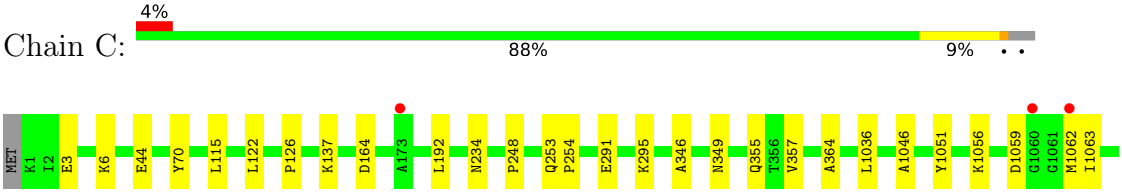
• Molecule 1: Maltose/maltodextrin-binding periplasmic protein,Putative pilin protein chimera

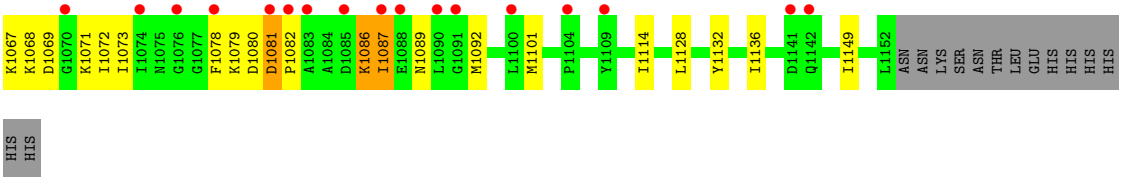
Chain B:



• Molecule 1: Maltose/maltodextrin-binding periplasmic protein,Putative pilin protein chimera

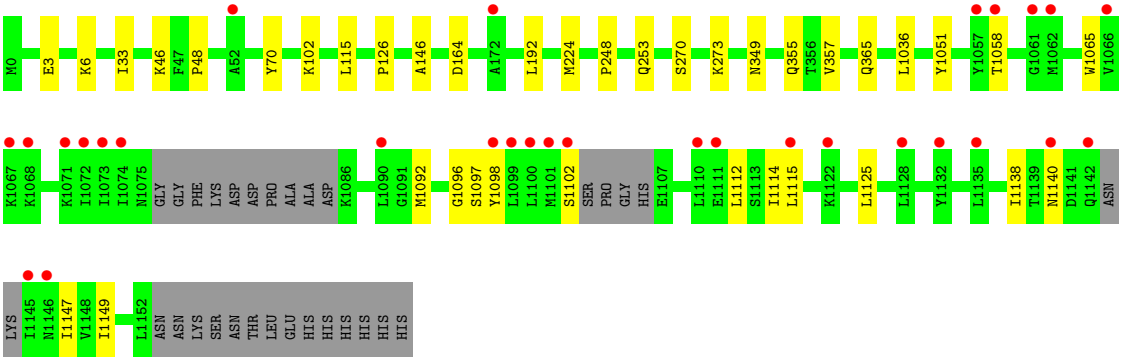
Chain C:





• Molecule 1: Maltose/maltodextrin-binding periplasmic protein,Putative pilin protein chimera

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	65.24Å 81.75Å 102.80Å 92.31° 91.05° 113.37°	Depositor
Resolution (Å)	39.70 – 2.30 39.70 – 2.30	Depositor EDS
% Data completeness (in resolution range)	81.7 (39.70-2.30) 81.7 (39.70-2.30)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.84 (at 2.29Å)	Xtriage
Refinement program	PHENIX (1.19.2_4158: ???)	Depositor
R, R_{free}	0.196 , 0.245 0.196 , 0.244	Depositor DCC
R_{free} test set	3522 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å^2)	28.4	Xtriage
Anisotropy	0.016	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 38.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.199 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	16068	wwPDB-VP
Average B, all atoms (Å^2)	52.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.04% 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	0.25	0/3820	0.45	0/5187
1	B	0.25	0/3905	0.43	0/5296
1	C	0.25	0/3917	0.44	0/5312
1	D	0.25	0/3800	0.45	0/5151
All	All	0.25	0/15442	0.44	0/20946

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	3731	0	3689	44	0
1	B	3810	0	3796	27	0
1	C	3824	0	3811	36	0
1	D	3708	0	3700	24	0
2	A	160	0	0	2	0
2	B	270	0	0	3	0
2	C	292	0	0	2	0
2	D	273	0	0	2	0
All	All	16068	0	14996	127	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.

The worst 5 of 127 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:91:PRO:HB3	1:C:1078:PHE:CE1	1.79	1.18
1:A:1108:LYS:HA	1:A:1152:LEU:HD22	1.53	0.89
1:A:1108:LYS:C	1:A:1152:LEU:HD21	2.01	0.79
1:C:1082:PRO:HD2	1:C:1086:LYS:HA	1.69	0.73
1:A:1109:TYR:N	1:A:1152:LEU:HD21	2.06	0.70

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	489/512 (96%)	458 (94%)	31 (6%)	0	100	100
1	B	499/512 (98%)	467 (94%)	30 (6%)	2 (0%)	34	42
1	C	498/512 (97%)	466 (94%)	30 (6%)	2 (0%)	34	42
1	D	479/512 (94%)	453 (95%)	25 (5%)	1 (0%)	47	58
All	All	1965/2048 (96%)	1844 (94%)	116 (6%)	5 (0%)	41	50

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	1081	ASP
1	C	1087	ILE
1	D	1140	ASN
1	B	1082	PRO
1	B	1081	ASP

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	380/409 (93%)	379 (100%)	1 (0%)	92 97
1	B	390/409 (95%)	390 (100%)	0	100 100
1	C	395/409 (97%)	392 (99%)	3 (1%)	81 91
1	D	381/409 (93%)	380 (100%)	1 (0%)	92 97
All	All	1546/1636 (94%)	1541 (100%)	5 (0%)	92 97

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1051	TYR
1	C	234	ASN
1	C	1051	TYR
1	C	1086	LYS
1	D	1051	TYR

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

Mol	Chain	Res	Type
1	A	49	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 monosaccharides 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	490/512 (95%)	0.07	32 (6%)	18	24	30, 58, 120, 205	0
1	B	497/512 (97%)	-0.20	30 (6%)	21	28	14, 36, 122, 194	0
1	C	497/512 (97%)	-0.27	20 (4%)	38	45	17, 38, 115, 177	0
1	D	482/512 (94%)	-0.24	30 (6%)	20	26	16, 37, 120, 176	0
All	All	1966/2048 (95%)	-0.16	112 (5%)	23	30	14, 43, 120, 205	0

The worst 5 of 112 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	1082	PRO	13.3
1	D	1115	LEU	10.5
1	B	1143	ASN	8.3
1	D	1102	SER	7.8
1	B	1068	LYS	7.7

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.