wwPDB X-ray Structure Validation Summary Report (i)

Mar 7, 2023 – 08:08 PM JST

PDB ID	:	8I5F
Title	:	Crystal structure of the DHR-2 domain of DOCK10 in complex with Cdc42 (T17N mutant)
Authors	:	Kukimoto-Niino, M.; Mishima-Tsumagari, C.; Fukui, Y.; Yokoyama, S.; Shirouzu, M.
Deposited on	:	2023-01-25
Resolution	:	2.80 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 (1)) were used in the production of this report:

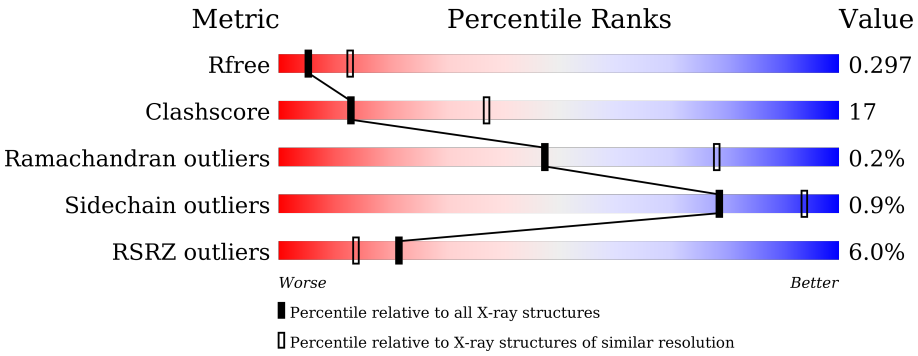
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.32.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.32.1

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 2.80 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	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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	494	4% 60% 26% • 12%
1	B	494	6% 54% 32% • 12%
2	C	195	% 64% 27% • 9%
2	D	195	12% 50% 39% • 9%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 9921 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 Dedicator of cytokinesis protein 10.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	433	Total	C	N	O	S	0	0	0
			3560	2279	593	670	18			
1	B	433	Total	C	N	O	S	0	0	0
			3560	2279	593	670	18			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1657	GLY	-	expression tag	UNP Q8BZN6
A	1658	SER	-	expression tag	UNP Q8BZN6
A	1659	SER	-	expression tag	UNP Q8BZN6
A	1660	GLY	-	expression tag	UNP Q8BZN6
A	1661	SER	-	expression tag	UNP Q8BZN6
A	1662	SER	-	expression tag	UNP Q8BZN6
A	1663	GLY	-	expression tag	UNP Q8BZN6
B	1657	GLY	-	expression tag	UNP Q8BZN6
B	1658	SER	-	expression tag	UNP Q8BZN6
B	1659	SER	-	expression tag	UNP Q8BZN6
B	1660	GLY	-	expression tag	UNP Q8BZN6
B	1661	SER	-	expression tag	UNP Q8BZN6
B	1662	SER	-	expression tag	UNP Q8BZN6
B	1663	GLY	-	expression tag	UNP Q8BZN6

- Molecule 2 is a protein called Cell division control protein 42 homolog.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	178	Total	C	N	O	S	0	0	0
			1389	893	222	267	7			
2	D	177	Total	C	N	O	S	0	0	0
			1380	888	221	264	7			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-6	GLY	-	expression tag	UNP P60953
C	-5	SER	-	expression tag	UNP P60953
C	-4	SER	-	expression tag	UNP P60953
C	-3	GLY	-	expression tag	UNP P60953
C	-2	SER	-	expression tag	UNP P60953
C	-1	SER	-	expression tag	UNP P60953
C	0	GLY	-	expression tag	UNP P60953
C	17	ASN	THR	engineered mutation	UNP P60953
C	188	SER	CYS	engineered mutation	UNP P60953
D	-6	GLY	-	expression tag	UNP P60953
D	-5	SER	-	expression tag	UNP P60953
D	-4	SER	-	expression tag	UNP P60953
D	-3	GLY	-	expression tag	UNP P60953
D	-2	SER	-	expression tag	UNP P60953
D	-1	SER	-	expression tag	UNP P60953
D	0	GLY	-	expression tag	UNP P60953
D	17	ASN	THR	engineered mutation	UNP P60953
D	188	SER	CYS	engineered mutation	UNP P60953

- Molecule 3 is water.

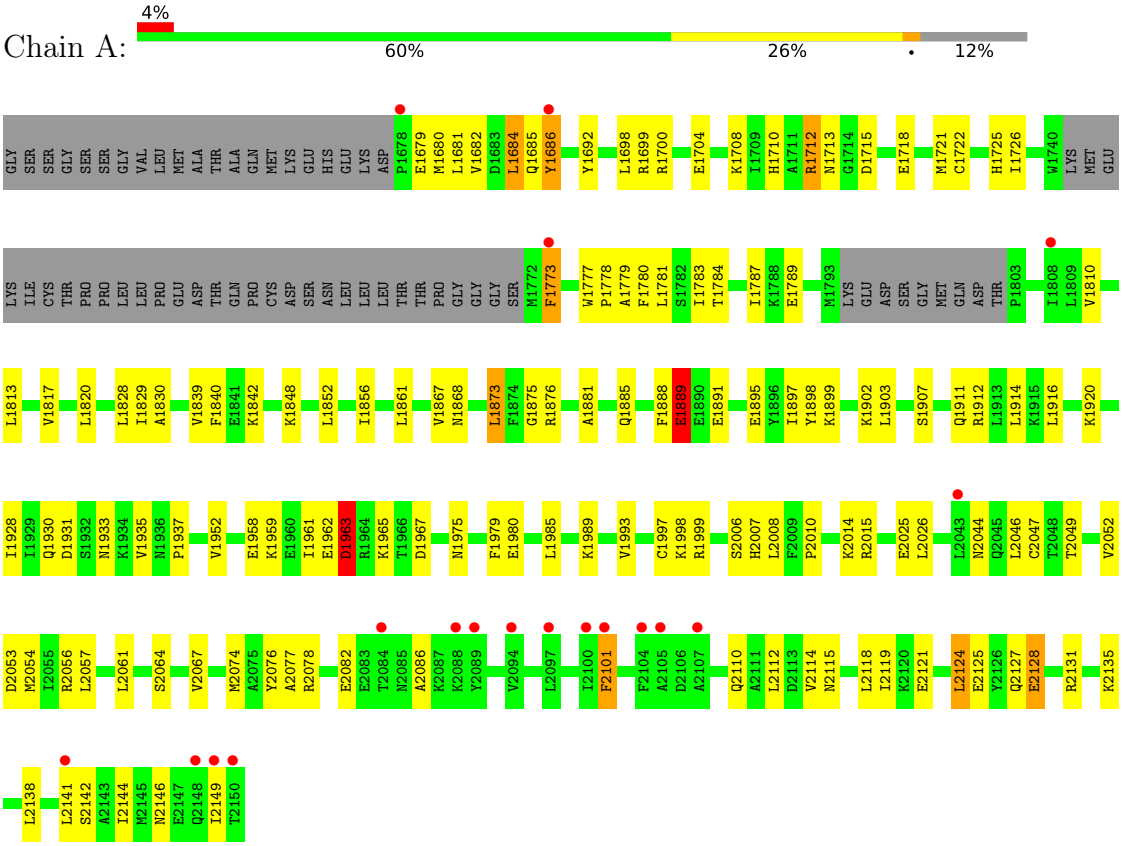
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	12	Total O 12 12	0	0
3	C	4	Total O 4 4	0	0
3	B	16	Total O 16 16	0	0

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.

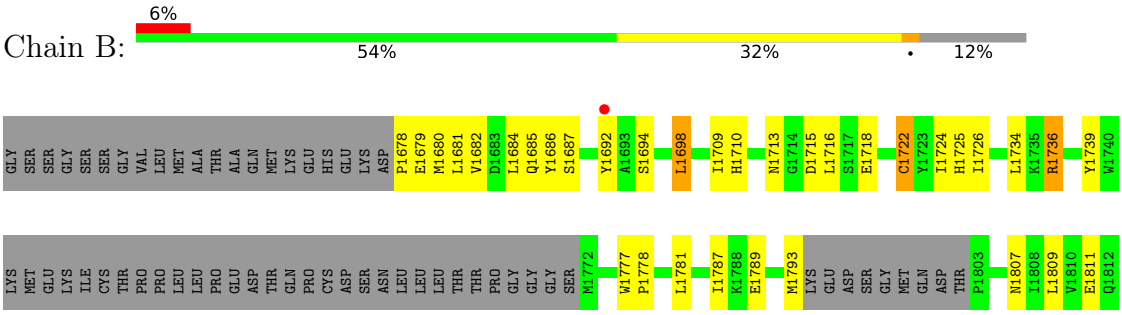
• Molecule 1: Dedicator of cytokinesis protein 10

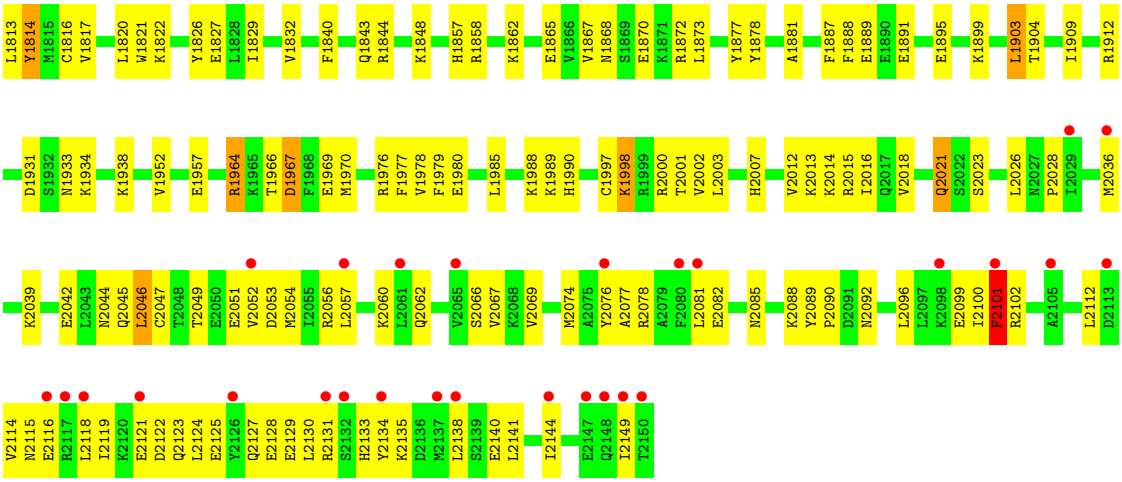
Chain A:



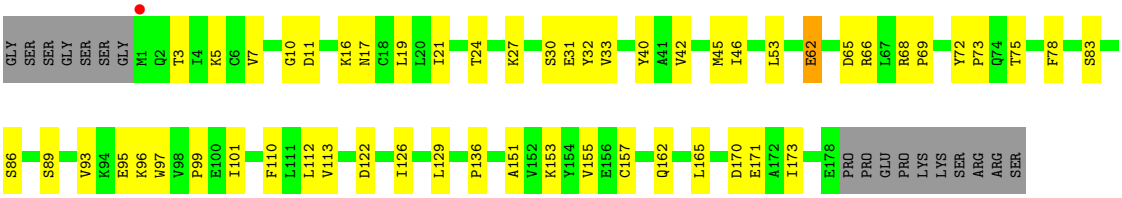
• Molecule 1: Dedicator of cytokinesis protein 10

Chain B:

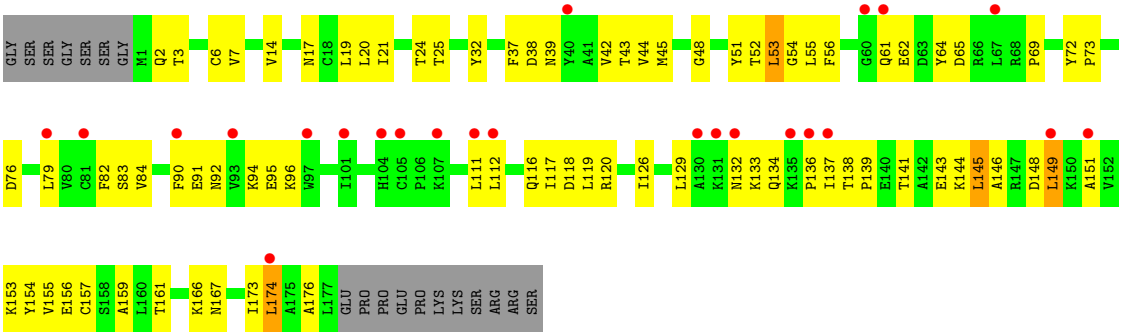




• Molecule 2: Cell division control protein 42 homolog



• Molecule 2: Cell division control protein 42 homolog



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	94.72Å 95.35Å 171.68Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.07 – 2.80 49.07 – 2.80	Depositor EDS
% Data completeness (in resolution range)	91.5 (49.07-2.80) 91.6 (49.07-2.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.85 (at 2.81Å)	Xtriage
Refinement program	PHENIX 1.18.2_3874	Depositor
R, R_{free}	0.226 , 0.296 0.226 , 0.297	Depositor DCC
R_{free} test set	3580 reflections (10.00%)	wwPDB-VP
Wilson B-factor (Å^2)	55.1	Xtriage
Anisotropy	0.643	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 48.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.000 for k,h,-l	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	9921	wwPDB-VP
Average B, all atoms (Å^2)	68.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 40.66 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 2.6494e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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	0.77	4/3633 (0.1%)	1.03	20/4891 (0.4%)
1	B	0.78	3/3633 (0.1%)	1.20	30/4891 (0.6%)
2	C	0.61	1/1419 (0.1%)	0.82	1/1931 (0.1%)
2	D	0.59	0/1410	1.20	8/1919 (0.4%)
All	All	0.73	8/10095 (0.1%)	1.09	59/13632 (0.4%)

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	A	0	2
1	B	0	2
All	All	0	4

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	1889	GLU	CD-OE2	9.46	1.36	1.25
2	C	62	GLU	CA-C	6.08	1.68	1.52
1	B	1722	CYS	CB-SG	5.82	1.92	1.82
1	A	1920	LYS	CE-NZ	5.79	1.63	1.49
1	B	1889	GLU	CD-OE1	5.60	1.31	1.25

The worst 5 of 59 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	D	145	LEU	CB-CG-CD2	-24.07	70.08	111.00
2	D	145	LEU	CB-CG-CD1	19.65	144.41	111.00
1	A	1985	LEU	CB-CG-CD2	-17.36	81.50	111.00
1	B	2101	PHE	CB-CG-CD2	-16.61	109.17	120.80
1	A	1985	LEU	CB-CG-CD1	14.46	135.57	111.00

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	1889	GLU	Sidechain
1	A	1963	ASP	Sidechain
1	B	2092	ASN	Sidechain
1	B	2101	PHE	Sidechain

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	3560	0	3541	112	1
1	B	3560	0	3541	126	1
2	C	1389	0	1404	37	0
2	D	1380	0	1398	71	0
3	A	12	0	0	6	0
3	B	16	0	0	7	0
3	C	4	0	0	1	0
All	All	9921	0	9884	327	1

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 17.

The worst 5 of 327 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:1684:LEU:HD13	1:B:1710:HIS:CE1	2.05	0.91
1:B:1873:LEU:HD13	1:B:1903:LEU:HD23	1.54	0.88
1:B:2123:GLN:HB3	2:D:73:PRO:HG2	1.58	0.86
2:D:2:GLN:HB3	2:D:51:TYR:CE1	2.13	0.83
2:D:14:VAL:HG13	2:D:83:SER:HB2	1.60	0.83

All (1) 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:A:1686:TYR:OH	1:B:1887:PHE:O[1_455]	2.11	0.09

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	427/494 (86%)	401 (94%)	26 (6%)	0	100 100
1	B	427/494 (86%)	401 (94%)	25 (6%)	1 (0%)	47 78
2	C	176/195 (90%)	167 (95%)	9 (5%)	0	100 100
2	D	175/195 (90%)	170 (97%)	4 (2%)	1 (1%)	25 56
All	All	1205/1378 (87%)	1139 (94%)	64 (5%)	2 (0%)	47 78

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	D	39	ASN
1	B	1903	LEU

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	388/440 (88%)	381 (98%)	7 (2%)	59 86
1	B	388/440 (88%)	385 (99%)	3 (1%)	81 94
2	C	158/172 (92%)	158 (100%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
2	D	157/172 (91%)	157 (100%)	0	100 100
All	All	1091/1224 (89%)	1081 (99%)	10 (1%)	78 94

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	1686	TYR
1	B	1857	HIS
1	B	1909	ILE
1	A	2049	THR
1	A	2064	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
2	D	17	ASN
1	B	2103	GLN
1	B	1725	HIS
2	C	132	ASN
1	B	1868	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 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	433/494 (87%)	0.17	19 (4%) 34 24	27, 51, 87, 105	0
1	B	433/494 (87%)	0.31	29 (6%) 17 10	34, 55, 112, 131	0
2	C	178/195 (91%)	-0.02	1 (0%) 89 86	41, 59, 95, 113	0
2	D	177/195 (90%)	0.86	24 (13%) 3 1	65, 121, 152, 174	0
All	All	1221/1378 (88%)	0.29	73 (5%) 21 14	27, 60, 131, 174	0

The worst 5 of 73 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	137	ILE	7.2
1	A	2101	PHE	5.8
1	B	2150	THR	5.7
2	D	136	PRO	5.0
1	B	2061	LEU	4.8

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.