wwPDB X-ray Structure Validation Summary Report (i)

Jun 18, 2024 – 10:33 PM EDT

PDB ID	:	4GYV
Title	:	Crystal structure of the DH domain of FARP2
Authors	:	He, X.; Zhang, X.
Deposited on	:	2012-09-05
Resolution	:	2.90 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:

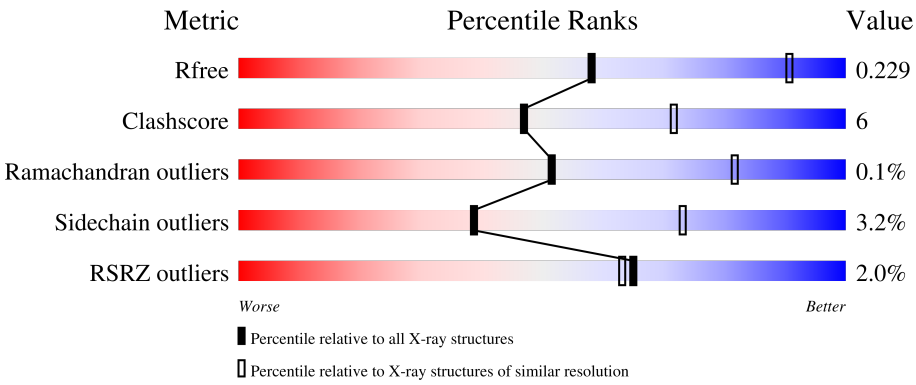
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.90 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	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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	218	83% 16% ••
1	B	218	83% 17% •
1	C	218	81% 17% •
1	D	218	% 86% 12% ••
1	E	218	6% 72% 26% ••

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	218	85% 12%
1	G	218	4% 76% 22%
1	H	218	86% 12%
1	I	218	5% 80% 18%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 15733 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 FERM, RhoGEF and pleckstrin domain-containing protein 2.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	216	1766	1134	301	323	4	4	0	0	0
1	B	216	1764	1131	302	323	4	4	0	0	0
1	C	216	1764	1131	302	323	4	4	0	0	0
1	D	216	1755	1123	303	321	4	4	0	0	0
1	E	215	1706	1096	289	313	4	4	0	0	0
1	F	216	1752	1122	299	323	4	4	0	0	0
1	G	215	1727	1108	297	314	4	4	0	0	0
1	H	216	1756	1125	300	323	4	4	0	0	0
1	I	215	1726	1107	295	316	4	4	0	0	0

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	532	GLY	-	EXPRESSION TAG	UNP Q91VS8
A	533	PRO	-	EXPRESSION TAG	UNP Q91VS8
A	534	HIS	-	EXPRESSION TAG	UNP Q91VS8
A	535	MSE	-	EXPRESSION TAG	UNP Q91VS8
B	532	GLY	-	EXPRESSION TAG	UNP Q91VS8
B	533	PRO	-	EXPRESSION TAG	UNP Q91VS8
B	534	HIS	-	EXPRESSION TAG	UNP Q91VS8
B	535	MSE	-	EXPRESSION TAG	UNP Q91VS8
C	532	GLY	-	EXPRESSION TAG	UNP Q91VS8
C	533	PRO	-	EXPRESSION TAG	UNP Q91VS8
C	534	HIS	-	EXPRESSION TAG	UNP Q91VS8

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	535	MSE	-	EXPRESSION TAG	UNP Q91VS8
D	532	GLY	-	EXPRESSION TAG	UNP Q91VS8
D	533	PRO	-	EXPRESSION TAG	UNP Q91VS8
D	534	HIS	-	EXPRESSION TAG	UNP Q91VS8
D	535	MSE	-	EXPRESSION TAG	UNP Q91VS8
E	532	GLY	-	EXPRESSION TAG	UNP Q91VS8
E	533	PRO	-	EXPRESSION TAG	UNP Q91VS8
E	534	HIS	-	EXPRESSION TAG	UNP Q91VS8
E	535	MSE	-	EXPRESSION TAG	UNP Q91VS8
F	532	GLY	-	EXPRESSION TAG	UNP Q91VS8
F	533	PRO	-	EXPRESSION TAG	UNP Q91VS8
F	534	HIS	-	EXPRESSION TAG	UNP Q91VS8
F	535	MSE	-	EXPRESSION TAG	UNP Q91VS8
G	532	GLY	-	EXPRESSION TAG	UNP Q91VS8
G	533	PRO	-	EXPRESSION TAG	UNP Q91VS8
G	534	HIS	-	EXPRESSION TAG	UNP Q91VS8
G	535	MSE	-	EXPRESSION TAG	UNP Q91VS8
H	532	GLY	-	EXPRESSION TAG	UNP Q91VS8
H	533	PRO	-	EXPRESSION TAG	UNP Q91VS8
H	534	HIS	-	EXPRESSION TAG	UNP Q91VS8
H	535	MSE	-	EXPRESSION TAG	UNP Q91VS8
I	532	GLY	-	EXPRESSION TAG	UNP Q91VS8
I	533	PRO	-	EXPRESSION TAG	UNP Q91VS8
I	534	HIS	-	EXPRESSION TAG	UNP Q91VS8
I	535	MSE	-	EXPRESSION TAG	UNP Q91VS8

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	4	4	4	0	0
2	B	3	3	3	0	0
2	C	3	3	3	0	0
2	D	1	1	1	0	0
2	F	3	3	3	0	0
2	G	1	1	1	0	0
2	H	1	1	1	0	0

Continued on next page...

Continued from previous page...

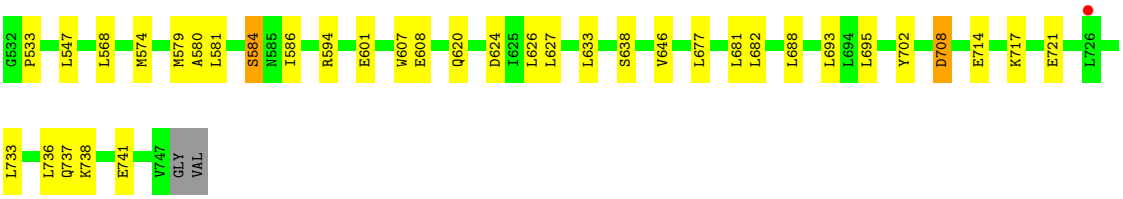
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	I	1	1	1	0	0

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.

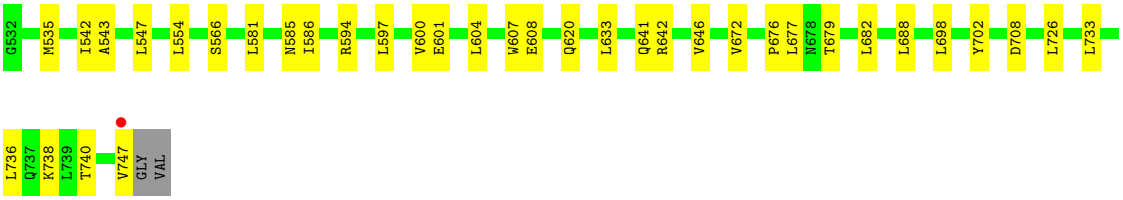
• Molecule 1: FERM, RhoGEF and pleckstrin domain-containing protein 2

Chain A:



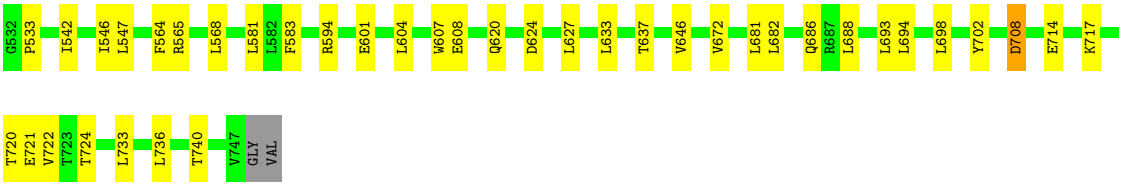
• Molecule 1: FERM, RhoGEF and pleckstrin domain-containing protein 2

Chain B:



• Molecule 1: FERM, RhoGEF and pleckstrin domain-containing protein 2

Chain C:

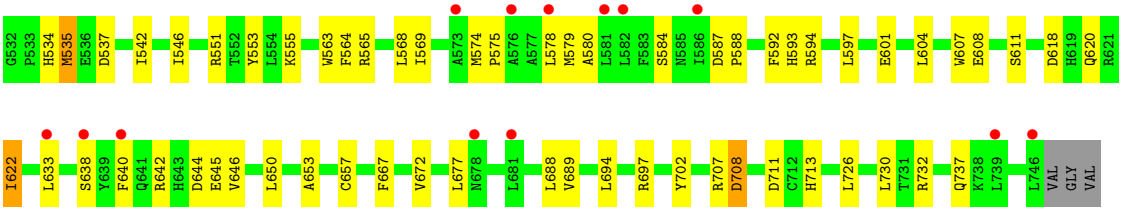


• Molecule 1: FERM, RhoGEF and pleckstrin domain-containing protein 2

Chain D:

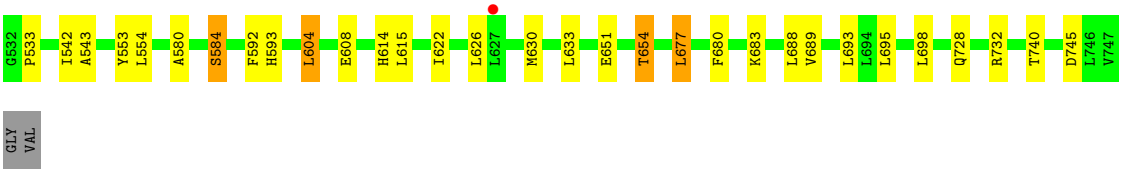
• Molecule 1: FERM, RhoGEF and pleckstrin domain-containing protein 2

Chain E:



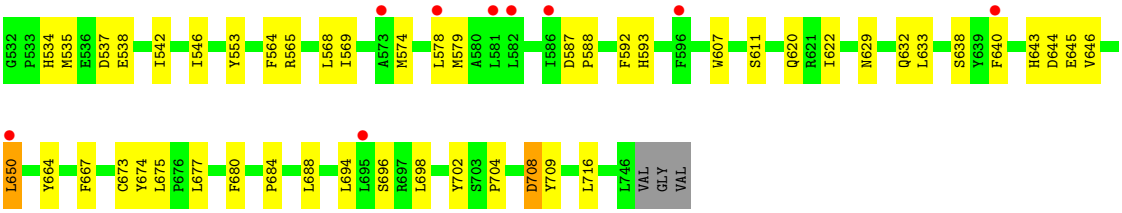
• Molecule 1: FERM, RhoGEF and pleckstrin domain-containing protein 2

Chain F:



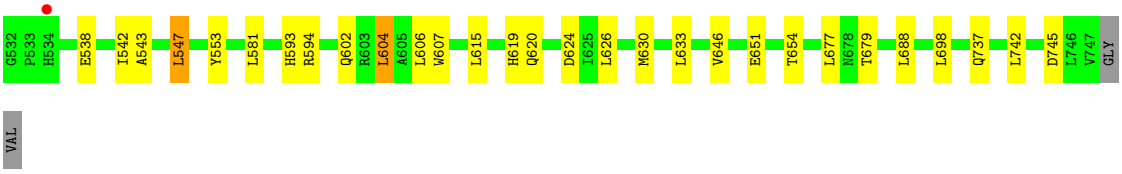
• Molecule 1: FERM, RhoGEF and pleckstrin domain-containing protein 2

Chain G:



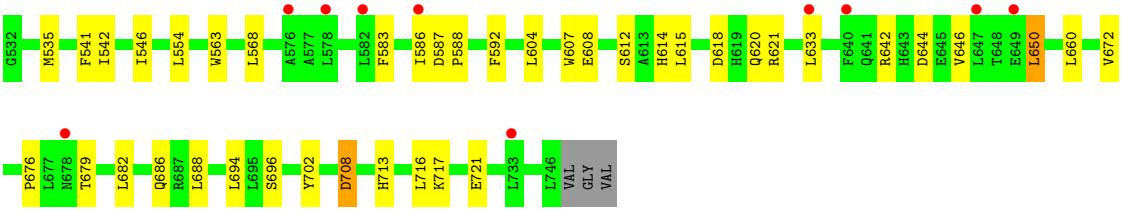
• Molecule 1: FERM, RhoGEF and pleckstrin domain-containing protein 2

Chain H:



• Molecule 1: FERM, RhoGEF and pleckstrin domain-containing protein 2

Chain I:


G532
M535
F541
I542
I546
L554
W563
L568
A576
A577
L578
L582
F583
I586
D587
P588
F592
L604
W607
E608
S612
A613
H614
L615
D618
H619
Q620
R621
L633
F640
Q641
R642
H643
D644
E645
V646
L647
T648
E649
L650
L660
V672
P676
L677
N678
T679
L682
Q686
R687
L688
L694
L695
S696
Y702
D708
H713
L716
K717
E721
L733
L746
VAL
GLY
VAL

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	121.19Å 209.99Å 325.01Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.58 – 2.90 48.58 – 2.90	Depositor EDS
% Data completeness (in resolution range)	96.6 (48.58-2.90) 96.6 (48.58-2.90)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.32 (at 2.91Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.178 , 0.221 0.182 , 0.229	Depositor DCC
R_{free} test set	1989 reflections (2.24%)	wwPDB-VP
Wilson B-factor (Å^2)	75.9	Xtriage
Anisotropy	0.076	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 51.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.487 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.487 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	15733	wwPDB-VP
Average B, all atoms (Å^2)	90.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.95% 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	0.53	0/1801	0.68	1/2433 (0.0%)
1	B	0.54	0/1799	0.66	1/2432 (0.0%)
1	C	0.54	0/1799	0.68	0/2432
1	D	0.46	0/1790	0.59	1/2422 (0.0%)
1	E	0.36	0/1740	0.53	0/2361
1	F	0.47	0/1787	0.60	1/2420 (0.0%)
1	G	0.35	0/1762	0.52	0/2387
1	H	0.45	0/1791	0.60	0/2424
1	I	0.37	0/1761	0.53	0/2385
All	All	0.46	0/16030	0.60	4/21696 (0.0%)

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	D	677	LEU	CA-CB-CG	-5.26	103.20	115.30
1	F	677	LEU	CA-CB-CG	-5.22	103.29	115.30
1	A	677	LEU	CA-CB-CG	-5.17	103.41	115.30
1	B	677	LEU	CA-CB-CG	-5.12	103.52	115.30

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	1766	0	1748	23	0
1	B	1764	0	1737	20	0
1	C	1764	0	1737	23	0
1	D	1755	0	1713	22	0
1	E	1706	0	1636	32	0
1	F	1752	0	1704	20	0
1	G	1727	0	1672	25	0
1	H	1756	0	1715	15	0
1	I	1726	0	1667	21	0
2	A	4	0	0	0	0
2	B	3	0	0	0	0
2	C	3	0	0	0	0
2	D	1	0	0	0	0
2	F	3	0	0	0	0
2	G	1	0	0	0	0
2	H	1	0	0	0	0
2	I	1	0	0	0	0
All	All	15733	0	15329	193	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 6.

The worst 5 of 193 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:G:574:MSE:HE2	1:G:578:LEU:HG	1.70	0.73
1:G:646:VAL:O	1:G:650:LEU:HB2	1.92	0.69
1:D:630:MSE:HE3	1:D:630:MSE:HA	1.74	0.69
1:F:580:ALA:O	1:F:584:SER:HB3	1.93	0.69
1:F:689:VAL:HG11	1:G:535:MSE:HG3	1.74	0.69

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	214/218 (98%)	209 (98%)	5 (2%)	0	100	100
1	B	214/218 (98%)	208 (97%)	6 (3%)	0	100	100
1	C	214/218 (98%)	205 (96%)	9 (4%)	0	100	100
1	D	214/218 (98%)	204 (95%)	10 (5%)	0	100	100
1	E	213/218 (98%)	196 (92%)	16 (8%)	1 (0%)	29	61
1	F	214/218 (98%)	204 (95%)	10 (5%)	0	100	100
1	G	213/218 (98%)	204 (96%)	9 (4%)	0	100	100
1	H	214/218 (98%)	208 (97%)	6 (3%)	0	100	100
1	I	213/218 (98%)	200 (94%)	12 (6%)	1 (0%)	29	61
All	All	1923/1962 (98%)	1838 (96%)	83 (4%)	2 (0%)	51	82

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	I	583	PHE
1	E	622	ILE

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	190/192 (99%)	186 (98%)	4 (2%)	53	81
1	B	189/192 (98%)	185 (98%)	4 (2%)	53	81
1	C	189/192 (98%)	185 (98%)	4 (2%)	53	81
1	D	186/192 (97%)	182 (98%)	4 (2%)	52	81
1	E	176/192 (92%)	164 (93%)	12 (7%)	16	42
1	F	186/192 (97%)	181 (97%)	5 (3%)	44	77
1	G	180/192 (94%)	173 (96%)	7 (4%)	32	66

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	H	187/192 (97%)	180 (96%)	7 (4%)	34	68
1	I	180/192 (94%)	173 (96%)	7 (4%)	32	66
All	All	1663/1728 (96%)	1609 (97%)	54 (3%)	39	73

5 of 54 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	584	SER
1	G	638	SER
1	I	612	SER
1	F	604	LEU
1	F	693	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	E	593	HIS
1	I	614	HIS

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 (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	212/218 (97%)	0.36	1 (0%)	91	91	42, 66, 97, 126	0
1	B	212/218 (97%)	0.33	1 (0%)	91	91	43, 67, 96, 124	0
1	C	212/218 (97%)	0.30	0	100	100	43, 68, 96, 127	0
1	D	212/218 (97%)	0.31	2 (0%)	84	84	56, 79, 133, 184	0
1	E	211/218 (96%)	0.46	13 (6%)	20	16	65, 117, 161, 197	0
1	F	212/218 (97%)	0.33	1 (0%)	91	91	58, 80, 133, 177	0
1	G	211/218 (96%)	0.48	9 (4%)	35	31	64, 117, 165, 191	0
1	H	212/218 (97%)	0.31	1 (0%)	91	91	59, 80, 130, 173	0
1	I	211/218 (96%)	0.42	10 (4%)	31	28	66, 119, 166, 196	0
All	All	1905/1962 (97%)	0.37	38 (1%)	65	63	42, 84, 152, 197	0

The worst 5 of 38 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	640	PHE	4.9
1	G	578	LEU	4.7
1	E	582	LEU	4.6
1	E	638	SER	4.5
1	G	586	ILE	3.9

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 monosaccharides 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.