wwPDB X-ray Structure Validation Summary Report (i)

Dec 2, 2024 – 06:29 PM EST

PDB ID	:	4BQ7
Title	:	Crystal structure of the RGMB-Neo1 complex form 2
Authors	:	Bell, C.H.; Healey, E.; van Erp, S.; Bishop, B.; Tang, C.; Gilbert, R.J.C.; Aricescu, A.R.; Pasterkamp, R.J.; Siebold, C.
Deposited on	:	2013-05-30
Resolution	:	6.60 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:

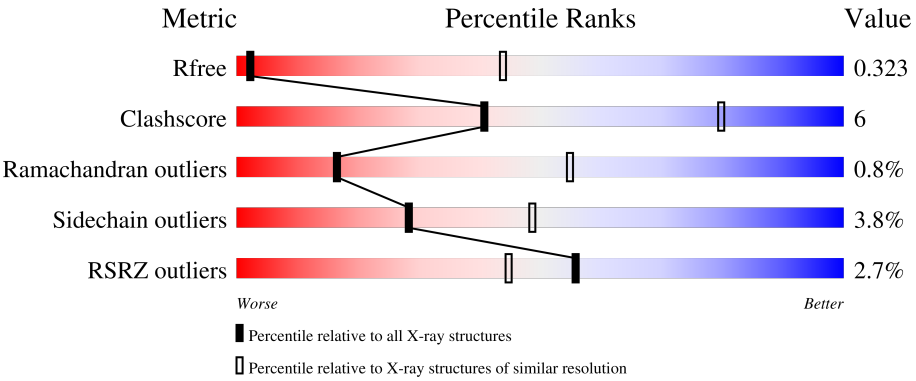
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.21
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.004 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.40

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 6.60 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}	164625	1100 (9.00-4.00)
Clashscore	180529	1140 (9.00-4.00)
Ramachandran outliers	177936	1010 (9.00-4.00)
Sidechain outliers	177891	1032 (9.00-3.96)
RSRZ outliers	164620	1095 (9.00-4.00)

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%

Mol	Chain	Length	Quality of chain
1	A	264	2% 64% 11% • 23%
1	B	264	2% 67% 8% • 23%
2	C	122	% 10% • • 88%
2	E	122	% 10% • • 88%
3	D	251	% 50% 8% 41%
3	F	251	2% 51% 7% 41%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 5714 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 NEOGENIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	202	Total	C	N	O	S	0	0	0
			1591	1017	270	298	6			
1	B	202	Total	C	N	O	S	0	0	0
			1591	1017	270	298	6			

There are 26 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	880	GLU	-	expression tag	UNP P97798
A	881	THR	-	expression tag	UNP P97798
A	882	GLY	-	expression tag	UNP P97798
A	1134	ASN	-	expression tag	UNP P97798
A	1135	GLY	-	expression tag	UNP P97798
A	1136	THR	-	expression tag	UNP P97798
A	1137	LYS	-	expression tag	UNP P97798
A	1138	HIS	-	expression tag	UNP P97798
A	1139	HIS	-	expression tag	UNP P97798
A	1140	HIS	-	expression tag	UNP P97798
A	1141	HIS	-	expression tag	UNP P97798
A	1142	HIS	-	expression tag	UNP P97798
A	1143	HIS	-	expression tag	UNP P97798
B	880	GLU	-	expression tag	UNP P97798
B	881	THR	-	expression tag	UNP P97798
B	882	GLY	-	expression tag	UNP P97798
B	1134	ASN	-	expression tag	UNP P97798
B	1135	GLY	-	expression tag	UNP P97798
B	1136	THR	-	expression tag	UNP P97798
B	1137	LYS	-	expression tag	UNP P97798
B	1138	HIS	-	expression tag	UNP P97798
B	1139	HIS	-	expression tag	UNP P97798
B	1140	HIS	-	expression tag	UNP P97798
B	1141	HIS	-	expression tag	UNP P97798
B	1142	HIS	-	expression tag	UNP P97798

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	1143	HIS	-	expression tag	UNP P97798

- Molecule 2 is a protein called RGM DOMAIN FAMILY MEMBER B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	15	Total	C	N	O	S	0	0	0
			118	79	16	21	2			
2	E	15	Total	C	N	O	S	0	0	0
			118	79	16	21	2			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	47	GLU	-	expression tag	UNP Q6NW40
C	48	THR	-	expression tag	UNP Q6NW40
C	49	GLY	-	expression tag	UNP Q6NW40
E	47	GLU	-	expression tag	UNP Q6NW40
E	48	THR	-	expression tag	UNP Q6NW40
E	49	GLY	-	expression tag	UNP Q6NW40

- Molecule 3 is a protein called RGM DOMAIN FAMILY MEMBER B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	148	Total	C	N	O	S	0	0	0
			1148	721	197	223	7			
3	F	148	Total	C	N	O	S	0	0	0
			1148	721	197	223	7			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	411	GLY	-	expression tag	UNP Q6NW40
D	412	THR	-	expression tag	UNP Q6NW40
D	413	LYS	-	expression tag	UNP Q6NW40
D	414	HIS	-	expression tag	UNP Q6NW40
D	415	HIS	-	expression tag	UNP Q6NW40
D	416	HIS	-	expression tag	UNP Q6NW40
D	417	HIS	-	expression tag	UNP Q6NW40
D	418	HIS	-	expression tag	UNP Q6NW40
D	419	HIS	-	expression tag	UNP Q6NW40
D	225	GLY	GLU	conflict	UNP Q6NW40
F	411	GLY	-	expression tag	UNP Q6NW40

Continued on next page...

Continued from previous page...

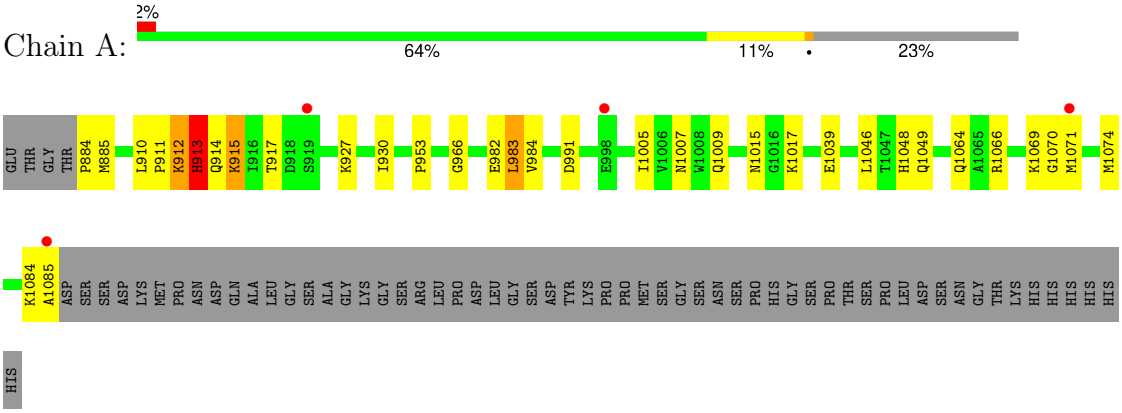
Chain	Residue	Modelled	Actual	Comment	Reference
F	412	THR	-	expression tag	UNP Q6NW40
F	413	LYS	-	expression tag	UNP Q6NW40
F	414	HIS	-	expression tag	UNP Q6NW40
F	415	HIS	-	expression tag	UNP Q6NW40
F	416	HIS	-	expression tag	UNP Q6NW40
F	417	HIS	-	expression tag	UNP Q6NW40
F	418	HIS	-	expression tag	UNP Q6NW40
F	419	HIS	-	expression tag	UNP Q6NW40
F	225	GLY	GLU	conflict	UNP Q6NW40

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. 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. 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.

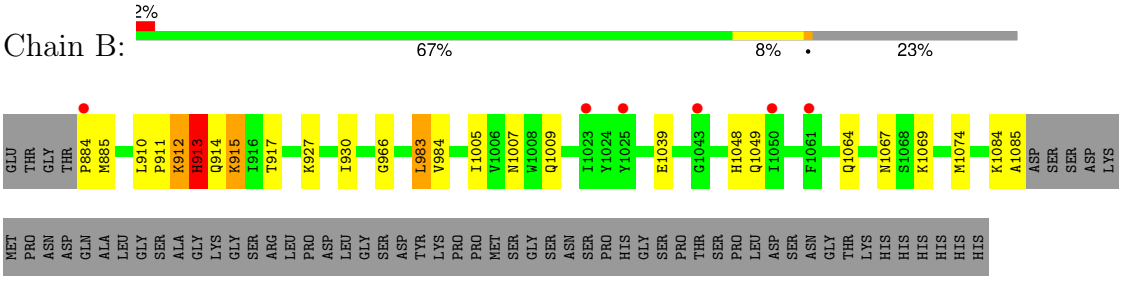
• Molecule 1: NEOGENIN

Chain A:



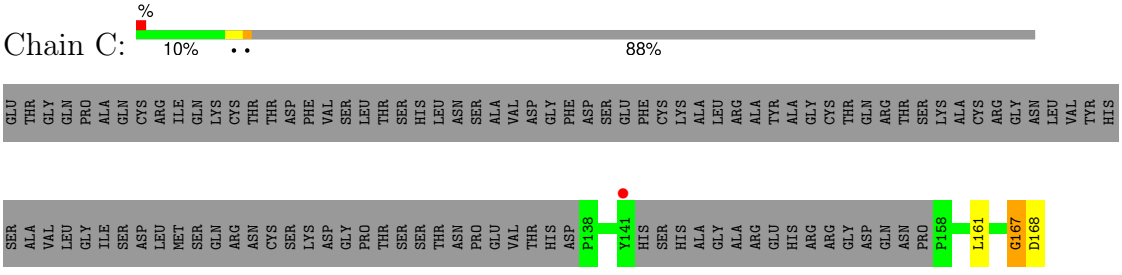
• Molecule 1: NEOGENIN

Chain B:



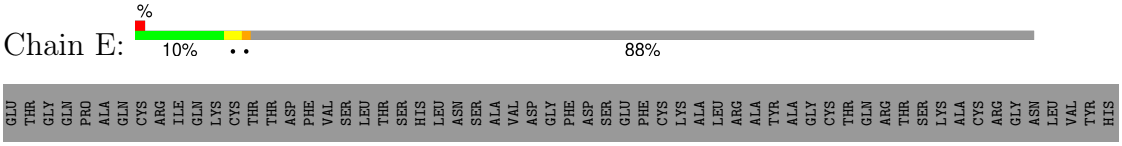
• Molecule 2: RGM DOMAIN FAMILY MEMBER B

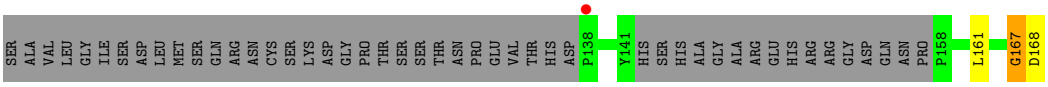
Chain C:



• Molecule 2: RGM DOMAIN FAMILY MEMBER B

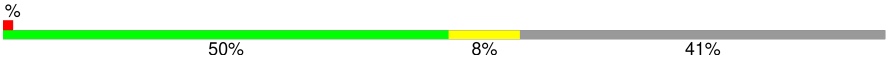
Chain E:

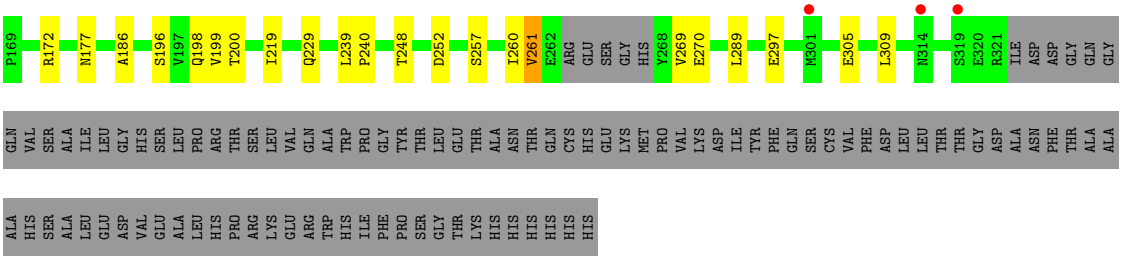




• Molecule 3: RGM DOMAIN FAMILY MEMBER B

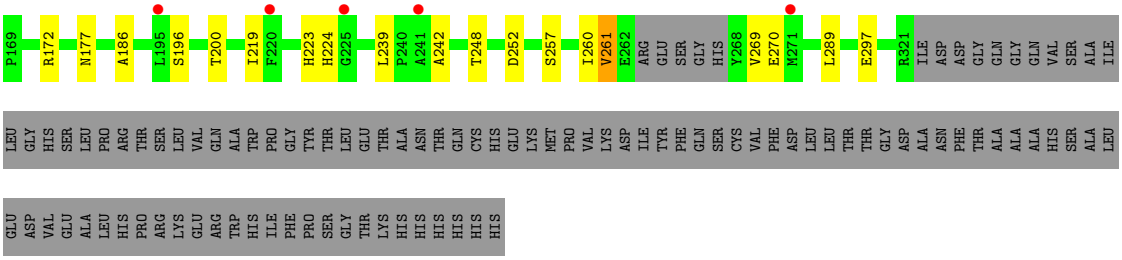
Chain D:





• Molecule 3: RGM DOMAIN FAMILY MEMBER B

Chain F:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32 1 2	Depositor
Cell constants a, b, c, α, β, γ	109.69Å 109.69Å 187.93Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	84.78 – 6.60 84.78 – 6.60	Depositor EDS
% Data completeness (in resolution range)	97.2 (84.78-6.60) 97.2 (84.78-6.60)	Depositor EDS
R_{merge}	0.23	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.72 (at 6.73Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, R_{free}	0.254 , 0.280 0.305 , 0.323	Depositor DCC
R_{free} test set	231 reflections (9.28%)	wwPDB-VP
Wilson B-factor (Å^2)	128.1	Xtriage
Anisotropy	0.133	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 186.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.42$, $< L^2 > = 0.25$	Xtriage
Estimated twinning fraction	0.077 for -h,-k,l	Xtriage
F_o,F_c correlation	0.75	EDS
Total number of atoms	5714	wwPDB-VP
Average B, all atoms (Å^2)	165.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.05% 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	0.68	1/1635 (0.1%)	0.80	4/2234 (0.2%)
1	B	0.82	1/1635 (0.1%)	0.71	1/2234 (0.0%)
2	C	0.46	0/122	1.01	1/162 (0.6%)
2	E	0.46	0/122	1.01	1/162 (0.6%)
3	D	0.51	0/1170	0.73	0/1590
3	F	0.51	0/1170	0.74	0/1590
All	All	0.65	2/5854 (0.0%)	0.76	7/7972 (0.1%)

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
2	C	0	1
2	E	0	1
All	All	0	2

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	983	LEU	C-N	26.18	1.94	1.34
1	A	983	LEU	C-N	-18.64	0.91	1.34

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	983	LEU	O-C-N	13.56	144.39	122.70
1	A	983	LEU	CA-C-N	-9.95	95.31	117.20
2	E	167	GLY	C-N-CA	8.68	143.40	121.70
2	C	167	GLY	C-N-CA	8.63	143.28	121.70
1	A	983	LEU	C-N-CA	-6.70	104.95	121.70

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	C	167	GLY	Peptide
2	E	167	GLY	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	1591	0	1585	46	0
1	B	1591	0	1585	16	0
2	C	118	0	101	1	0
2	E	118	0	101	1	0
3	D	1148	0	1122	33	1
3	F	1148	0	1122	14	0
All	All	5714	0	5616	72	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 6.

The worst 5 of 72 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:983:LEU:C	1:B:984:VAL:N	1.94	1.21
1:A:1009:GLN:HG3	3:D:200:THR:OG1	1.68	0.94
1:A:991:ASP:OD2	3:D:200:THR:HG21	1.68	0.93
1:A:1009:GLN:HG3	3:D:200:THR:HG1	1.35	0.92
1:A:1009:GLN:HG2	3:D:186:ALA:HB2	1.55	0.86

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)
3:D:305:GLU:OE1	3:D:305:GLU:OE1[4_445]	1.91	0.29

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	200/264 (76%)	187 (94%)	10 (5%)	3 (2%)	8 40
1	B	200/264 (76%)	187 (94%)	10 (5%)	3 (2%)	8 40
2	C	11/122 (9%)	11 (100%)	0	0	100 100
2	E	11/122 (9%)	11 (100%)	0	0	100 100
3	D	144/251 (57%)	134 (93%)	10 (7%)	0	100 100
3	F	144/251 (57%)	133 (92%)	11 (8%)	0	100 100
All	All	710/1274 (56%)	663 (93%)	41 (6%)	6 (1%)	16 55

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	912	LYS
1	A	913	HIS
1	B	912	LYS
1	B	913	HIS
1	B	966	GLY

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	178/232 (77%)	173 (97%)	5 (3%)	38 57
1	B	178/232 (77%)	173 (97%)	5 (3%)	38 57
2	C	13/103 (13%)	12 (92%)	1 (8%)	10 30

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	E	13/103 (13%)	12 (92%)	1 (8%)	10	30
3	D	128/215 (60%)	122 (95%)	6 (5%)	22	44
3	F	128/215 (60%)	122 (95%)	6 (5%)	22	44
All	All	638/1100 (58%)	614 (96%)	24 (4%)	28	49

5 of 24 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	257	SER
2	E	168	ASP
3	D	269	VAL
3	F	172	ARG
1	B	913	HIS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	1009	GLN
3	D	229	GLN
3	F	224	HIS

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

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

5.5 Carbohydrates

There are no oligosaccharides in this entry.

5.6 Ligand geometry

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)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	B	1
1	A	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	B	983:LEU	C	984:VAL	N	1.94
1	A	983:LEU	C	984:VAL	N	0.91

6 Fit of model and data

6.1 Protein, DNA and RNA chains

Warning: The R factor obtained from EDS is 0.318, which does not match the depositor's R factor of 0.2537. Please interpret the results in this section carefully.

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	202/264 (76%)	0.20	4 (1%)	64	53	132, 150, 189, 225	0
1	B	202/264 (76%)	0.15	6 (2%)	52	42	132, 177, 214, 238	0
2	C	15/122 (12%)	0.37	1 (6%)	25	24	140, 153, 176, 188	0
2	E	15/122 (12%)	0.34	1 (6%)	25	24	157, 164, 172, 174	0
3	D	148/251 (58%)	0.17	3 (2%)	64	53	134, 150, 174, 203	0
3	F	148/251 (58%)	0.30	5 (3%)	48	39	154, 167, 182, 209	0
All	All	730/1274 (57%)	0.20	20 (2%)	56	45	132, 164, 203, 238	0

The worst 5 of 20 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	E	138	PRO	3.3
1	A	919	SER	3.1
1	B	1061	PHE	3.0
1	A	1085	ALA	2.8
1	B	1025	TYR	2.7

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

There are no such residues in this entry.