



# wwPDB X-ray Structure Validation Summary Report ⓘ

Apr 29, 2024 – 08:33 AM EDT

PDB ID : 8TH5  
Title : Crystal Structure of the G3BP1 NTF2-like domain bound to the IDR1 of SARS-CoV-2 nucleocapsid protein P13L mutant  
Authors : Hughes, M.P.; Taylor, J.P.; Yang, Z.  
Deposited on : 2023-07-13  
Resolution : 2.62 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](mailto: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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Xtriage (Phenix) : 1.13  
EDS : 2.36.2  
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) : 2.36.2

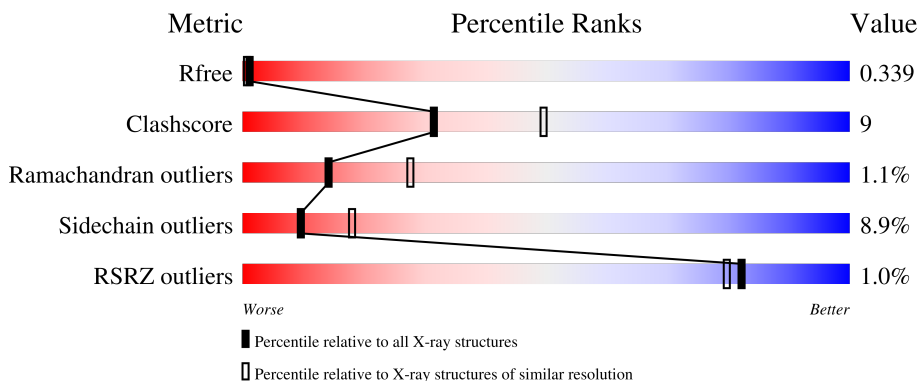
# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.62 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	3797 (2.64-2.60)
Clashscore	141614	4168 (2.64-2.60)
Ramachandran outliers	138981	4093 (2.64-2.60)
Sidechain outliers	138945	4093 (2.64-2.60)
RSRZ outliers	127900	3731 (2.64-2.60)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	139	 65% 20% 12%
1	B	139	 66% 20% 12%
1	C	139	 67% 22% 10%
1	D	139	 76% 17% 7%
1	E	139	 71% 23% 6%

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	139	 % 70% 19% • • 6%
1	G	139	 72% 15% • 10%
1	H	139	 2% 63% 24% 5% 7%
1	I	139	 68% 15% • 14%
1	J	139	 4% 53% 17% • 27%
2	K	25	 28% 8% 64%
2	L	25	 32% 68%
2	M	25	 20% 12% • 64%
2	N	25	 20% 12% • 64%
2	O	25	 28% • • 64%
2	P	25	 20% • 76%
2	Q	25	 24% 8% 68%

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 10350 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 Ras GTPase-activating protein-binding protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	123	1007	639	184	179	5	0	1	0
1	B	122	995	633	180	177	5	0	1	0
1	C	125	1023	650	186	182	5	0	1	0
1	D	129	1045	664	187	187	7	0	0	0
1	E	131	1055	670	192	189	4	0	0	0
1	F	130	1063	677	195	186	5	0	2	0
1	G	125	990	636	177	172	5	0	0	0
1	H	129	1013	644	182	183	4	0	0	0
1	I	119	905	578	162	161	4	0	1	0
1	J	102	798	512	141	141	4	0	1	0

- Molecule 2 is a protein called Nucleoprotein.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
2	K	9	65	43	12	10	0	0	0
2	P	6	41	28	6	7	0	0	0
2	L	8	51	34	8	9	0	0	0
2	M	9	62	40	12	10	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace	
2	O	9	Total	C	N	O	0	0	0
			59	40	9	10			
2	Q	8	Total	C	N	O	0	0	0
			60	40	11	9			
2	N	9	Total	C	N	O	0	0	0
			54	35	9	10			

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
K	13	LEU	PRO	engineered mutation	UNP P0DTC9
P	13	LEU	PRO	engineered mutation	UNP P0DTC9
L	13	LEU	PRO	engineered mutation	UNP P0DTC9
M	13	LEU	PRO	engineered mutation	UNP P0DTC9
O	13	LEU	PRO	engineered mutation	UNP P0DTC9
Q	13	LEU	PRO	engineered mutation	UNP P0DTC9
N	13	LEU	PRO	engineered mutation	UNP P0DTC9

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	6	Total	O	0	0
			6	6		
3	B	5	Total	O	0	0
			5	5		
3	C	12	Total	O	0	0
			12	12		
3	D	3	Total	O	0	0
			3	3		
3	E	7	Total	O	0	0
			7	7		
3	F	5	Total	O	0	0
			5	5		
3	G	8	Total	O	0	0
			8	8		
3	H	6	Total	O	0	0
			6	6		
3	I	8	Total	O	0	0
			8	8		
3	J	2	Total	O	0	0
			2	2		
3	P	1	Total	O	0	0
			1	1		

*Continued on next page...*

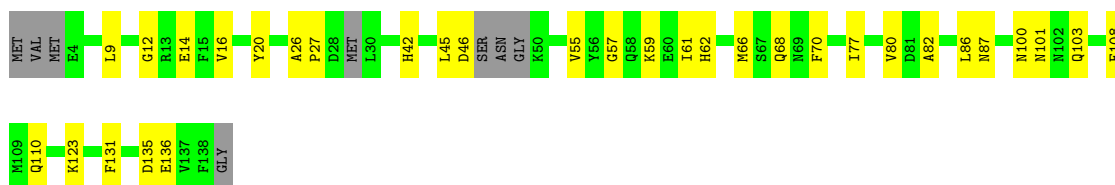
*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
3	O	1	Total	O	0	0
			1	1		



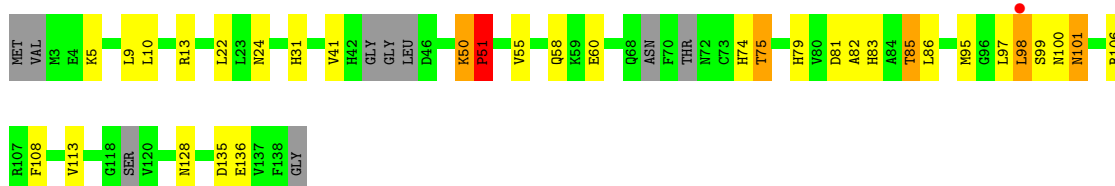
- Molecule 1: Ras GTPase-activating protein-binding protein 1

Chain E: 



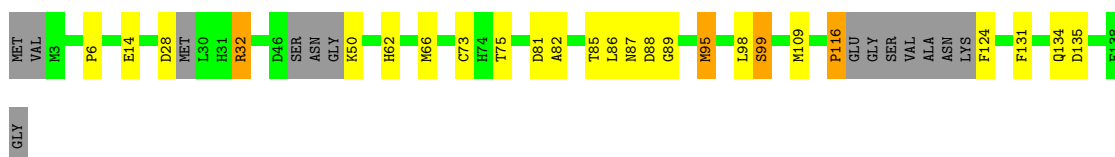
- Molecule 1: Ras GTPase-activating protein-binding protein 1

Chain F: 



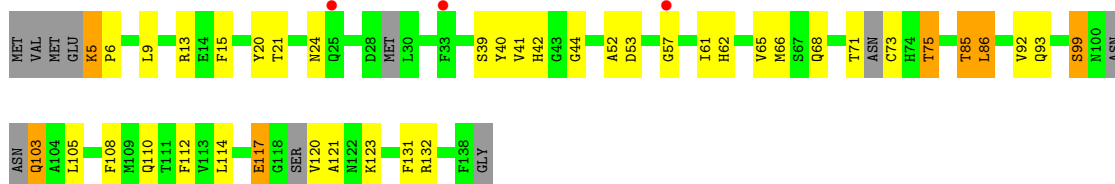
- Molecule 1: Ras GTPase-activating protein-binding protein 1

Chain G: 



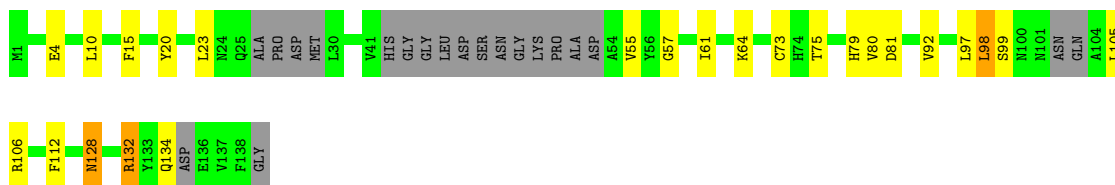
- Molecule 1: Ras GTPase-activating protein-binding protein 1

Chain H: 



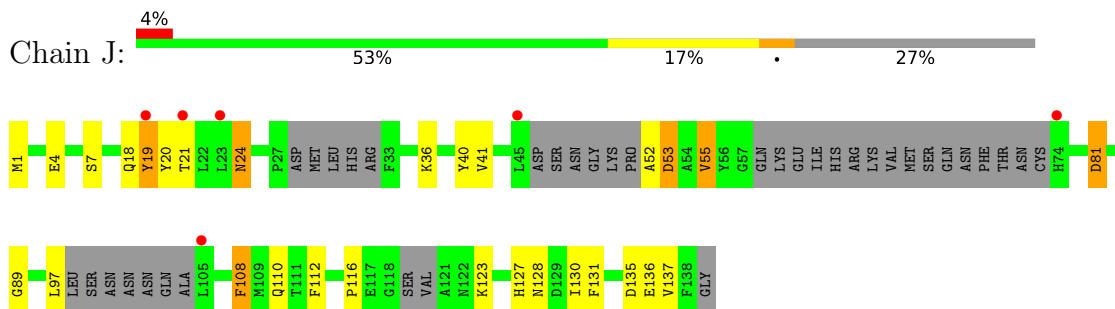
- Molecule 1: Ras GTPase-activating protein-binding protein 1

Chain I: 

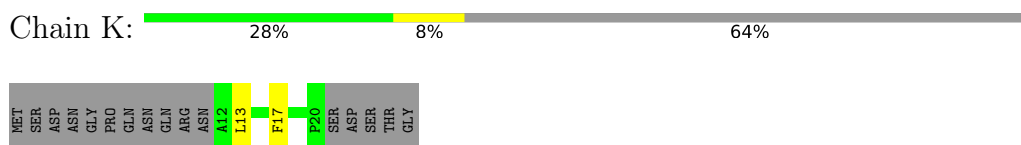




- Molecule 1: Ras GTPase-activating protein-binding protein 1



- Molecule 2: Nucleoprotein



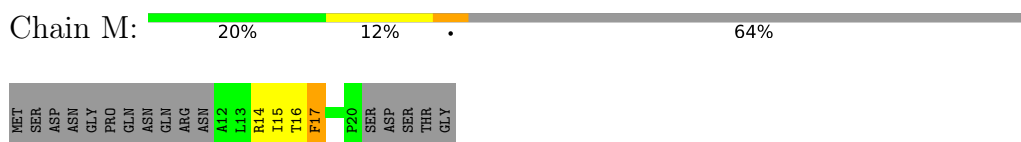
- Molecule 2: Nucleoprotein



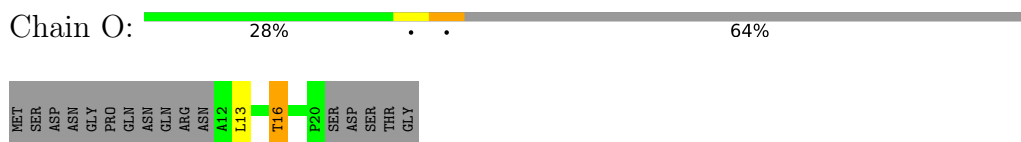
- Molecule 2: Nucleoprotein



- Molecule 2: Nucleoprotein



- Molecule 2: Nucleoprotein

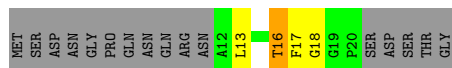


- Molecule 2: Nucleoprotein





- Molecule 2: Nucleoprotein



## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	61.60Å 69.45Å 101.04Å 70.66° 76.86° 88.31°	Depositor
Resolution (Å)	46.70 – 2.62 46.65 – 2.62	Depositor EDS
% Data completeness (in resolution range)	71.4 (46.70-2.62) 71.4 (46.65-2.62)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.91 (at 2.61Å)	Xtrriage
Refinement program	REFMAC 5.8.0419	Depositor
R, $R_{free}$	0.226 , 0.336 0.229 , 0.339	Depositor DCC
$R_{free}$ test set	1633 reflections (4.97%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	37.9	Xtrriage
Anisotropy	0.073	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 21.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.050 for -h,k,k-l	Xtrriage
$F_o, F_c$ correlation	0.89	EDS
Total number of atoms	10350	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	36.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.87% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  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	0.33	0/1032	0.75	4/1397 (0.3%)
1	B	0.32	0/1019	0.69	0/1375
1	C	0.30	0/1048	0.66	0/1418
1	D	0.32	0/1069	0.61	0/1445
1	E	0.32	0/1079	0.65	0/1457
1	F	0.33	0/1088	0.64	0/1465
1	G	0.32	0/1012	0.63	0/1367
1	H	0.33	0/1035	0.66	0/1398
1	I	0.34	0/924	0.66	0/1251
1	J	0.35	0/813	0.70	0/1094
2	K	0.43	0/66	0.69	0/88
2	L	0.43	0/52	0.50	0/70
2	M	0.34	0/63	0.70	0/84
2	N	0.36	0/54	0.65	0/72
2	O	0.38	0/60	0.63	0/81
2	P	0.30	0/41	0.58	0/55
2	Q	0.42	0/61	0.68	0/81
All	All	0.33	0/10516	0.66	4/14198 (0.0%)

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	G	0	1
1	I	0	1
All	All	0	2

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	20[A]	TYR	N-CA-CB	6.54	122.37	110.60
1	A	20[B]	TYR	N-CA-CB	6.54	122.37	110.60
1	A	20[A]	TYR	CB-CA-C	-5.28	99.83	110.40
1	A	20[B]	TYR	CB-CA-C	-5.28	99.83	110.40

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	G	32	ARG	Sidechain
1	I	132	ARG	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	1007	0	967	28	0
1	B	995	0	975	17	0
1	C	1023	0	981	18	0
1	D	1045	0	1011	15	0
1	E	1055	0	1014	18	0
1	F	1063	0	1025	23	0
1	G	990	0	948	14	0
1	H	1013	0	951	23	0
1	I	905	0	834	12	0
1	J	798	0	751	16	0
2	K	65	0	68	1	0
2	L	51	0	43	0	0
2	M	62	0	59	3	0
2	N	54	0	42	3	0
2	O	59	0	57	2	0
2	P	41	0	35	1	0
2	Q	60	0	63	1	0
3	A	6	0	0	0	0
3	B	5	0	0	0	0
3	C	12	0	0	0	0
3	D	3	0	0	0	0
3	E	7	0	0	1	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	F	5	0	0	0	0
3	G	8	0	0	0	0
3	H	6	0	0	0	0
3	I	8	0	0	0	0
3	J	2	0	0	0	0
3	O	1	0	0	0	0
3	P	1	0	0	0	0
All	All	10350	0	9824	181	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 9.

The worst 5 of 181 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:J:19:TYR:O	1:J:21:THR:N	1.85	1.07
1:E:135:ASP:OD1	1:E:136:GLU:N	2.11	0.83
1:F:9:LEU:HA	1:F:82:ALA:HB3	1.66	0.77
1:F:50:LYS:CB	1:F:51:PRO:HD2	2.19	0.73
1:A:20[B]:TYR:CD2	1:A:77:ILE:HD11	2.26	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	120/139 (86%)	107 (89%)	13 (11%)	0	100	100
1	B	117/139 (84%)	107 (92%)	10 (8%)	0	100	100
1	C	122/139 (88%)	111 (91%)	11 (9%)	0	100	100
1	D	125/139 (90%)	121 (97%)	4 (3%)	0	100	100

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	E	125/139 (90%)	115 (92%)	9 (7%)	1 (1%)	19	36
1	F	123/139 (88%)	107 (87%)	13 (11%)	3 (2%)	6	9
1	G	117/139 (84%)	99 (85%)	18 (15%)	0	100	100
1	H	119/139 (86%)	111 (93%)	7 (6%)	1 (1%)	19	36
1	I	110/139 (79%)	96 (87%)	13 (12%)	1 (1%)	17	33
1	J	91/139 (66%)	76 (84%)	9 (10%)	6 (7%)	1	1
2	K	7/25 (28%)	7 (100%)	0	0	100	100
2	L	6/25 (24%)	6 (100%)	0	0	100	100
2	M	7/25 (28%)	5 (71%)	1 (14%)	1 (14%)	0	0
2	N	7/25 (28%)	3 (43%)	3 (43%)	1 (14%)	0	0
2	O	7/25 (28%)	7 (100%)	0	0	100	100
2	P	4/25 (16%)	4 (100%)	0	0	100	100
2	Q	6/25 (24%)	6 (100%)	0	0	100	100
All	All	1213/1565 (78%)	1088 (90%)	111 (9%)	14 (1%)	14	25

5 of 14 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	J	19	TYR
1	J	20[A]	TYR
1	J	20[B]	TYR
1	J	137	VAL
2	N	16	THR

### 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	108/121 (89%)	98 (91%)	10 (9%)	9	16
1	B	108/121 (89%)	98 (91%)	10 (9%)	9	16
1	C	110/121 (91%)	102 (93%)	8 (7%)	14	27

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	D	113/121 (93%)	108 (96%)	5 (4%)	28	52
1	E	112/121 (93%)	107 (96%)	5 (4%)	27	50
1	F	112/121 (93%)	102 (91%)	10 (9%)	9	18
1	G	102/121 (84%)	92 (90%)	10 (10%)	8	14
1	H	105/121 (87%)	89 (85%)	16 (15%)	3	4
1	I	89/121 (74%)	83 (93%)	6 (7%)	16	31
1	J	79/121 (65%)	69 (87%)	10 (13%)	4	7
2	K	6/20 (30%)	5 (83%)	1 (17%)	2	3
2	L	4/20 (20%)	4 (100%)	0	100	100
2	M	5/20 (25%)	3 (60%)	2 (40%)	0	0
2	N	3/20 (15%)	1 (33%)	2 (67%)	0	0
2	O	5/20 (25%)	4 (80%)	1 (20%)	1	2
2	P	3/20 (15%)	2 (67%)	1 (33%)	0	0
2	Q	6/20 (30%)	5 (83%)	1 (17%)	2	3
All	All	1070/1350 (79%)	972 (91%)	98 (9%)	9	16

5 of 98 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	135	ASP
1	H	117	GLU
1	H	21	THR
1	H	75	THR
1	I	81	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 36 such sidechains are listed below:

Mol	Chain	Res	Type
1	G	128	ASN
1	J	127	HIS
1	G	134	GLN
1	I	18	GLN
1	C	127	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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	123/139 (88%)	-0.47	2 (1%) 72 68	7, 23, 38, 66	0
1	B	122/139 (87%)	-0.43	1 (0%) 86 84	10, 28, 52, 72	1 (0%)
1	C	125/139 (89%)	-0.51	0 100 100	13, 29, 50, 69	0
1	D	129/139 (92%)	-0.42	0 100 100	14, 30, 53, 58	0
1	E	131/139 (94%)	-0.39	0 100 100	12, 29, 57, 68	0
1	F	130/139 (93%)	-0.30	1 (0%) 86 84	14, 38, 61, 73	1 (0%)
1	G	125/139 (89%)	-0.46	0 100 100	17, 37, 56, 60	0
1	H	129/139 (92%)	-0.13	3 (2%) 60 55	22, 43, 64, 76	0
1	I	119/139 (85%)	0.02	0 100 100	20, 48, 75, 96	1 (0%)
1	J	102/139 (73%)	0.18	6 (5%) 22 17	23, 51, 77, 88	0
2	K	9/25 (36%)	-0.38	0 100 100	18, 25, 50, 57	0
2	L	8/25 (32%)	-0.34	0 100 100	28, 30, 44, 51	0
2	M	9/25 (36%)	-0.28	0 100 100	33, 48, 53, 56	0
2	N	9/25 (36%)	0.15	0 100 100	35, 51, 65, 66	0
2	O	9/25 (36%)	-0.11	0 100 100	37, 42, 48, 49	0
2	P	6/25 (24%)	-0.50	0 100 100	35, 38, 40, 42	0
2	Q	8/25 (32%)	-0.56	0 100 100	39, 56, 63, 65	0
All	All	1293/1565 (82%)	-0.30	13 (1%) 82 80	7, 35, 63, 96	3 (0%)

The worst 5 of 13 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	51	PRO	4.0
1	B	98[A]	LEU	3.4
1	J	21	THR	3.2
1	J	74	HIS	3.2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	J	23	LEU	3.0

## 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.