



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 03:32 am BST

PDB ID : 3QRF  
Title : Structure of a domain-swapped FOXP3 dimer  
Authors : Bandukwala, H.S.; Wu, Y.; Feurer, M.; Chen, Y.; Barbosa, B.; Ghosh, S.;  
Stroud, J.C.; Benoist, C.; Mathis, D.; Rao, A.; Chen, L.  
Deposited on : 2011-02-17  
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](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 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	:	2.11
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)	:	2.11

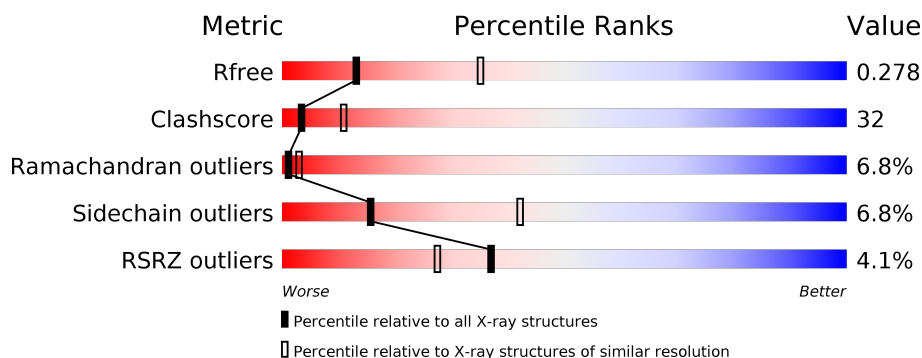
# 1 Overall quality at a glance

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.




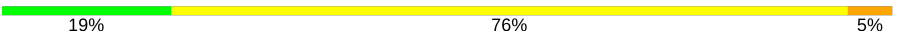
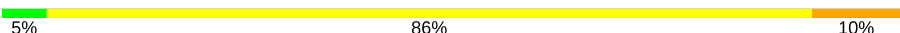
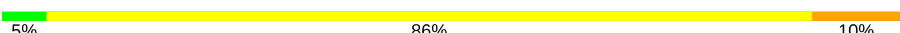
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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 on 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	M	286	<div> <div>6%</div> <div> <div>41%</div> <div>49%</div> <div>9%</div> </div> </div>
1	N	286	<div> <div>5%</div> <div> <div>43%</div> <div>47%</div> <div>9%</div> </div> </div>
2	F	82	<div> <div>%</div> <div> <div>39%</div> <div>54%</div> <div>7%</div> </div> </div>
2	G	82	<div> <div>5%</div> <div> <div>57%</div> <div>39%</div> <div>•</div> </div> </div>
2	H	82	<div> <div>2%</div> <div> <div>44%</div> <div>49%</div> <div>7%</div> </div> </div>
2	I	82	<div> <div>2%</div> <div> <div>57%</div> <div>39%</div> <div>•</div> </div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	A	21	 24%71%5%
3	C	21	 19%76%5%
4	B	21	 5%86%10%
4	D	21	 5%86%10%

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 9105 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 Nuclear factor of activated T-cells, cytoplasmic 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	N	286	Total	C	N	O	S	0	0	0
			2264	1421	413	421	9			
1	M	286	Total	C	N	O	S	0	0	0
			2264	1421	413	421	9			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
N	393	SER	-	EXPRESSION TAG	UNP Q13469
N	394	SER	-	EXPRESSION TAG	UNP Q13469
N	395	VAL	-	EXPRESSION TAG	UNP Q13469
N	629	ALA	ASP	CONFLICT	UNP Q13469
N	630	ALA	LYS	CONFLICT	UNP Q13469
N	631	ALA	ASP	CONFLICT	UNP Q13469
N	632	ALA	LYS	CONFLICT	UNP Q13469
N	633	ALA	SER	CONFLICT	UNP Q13469
N	634	ALA	GLN	CONFLICT	UNP Q13469
M	393	SER	-	EXPRESSION TAG	UNP Q13469
M	394	SER	-	EXPRESSION TAG	UNP Q13469
M	395	VAL	-	EXPRESSION TAG	UNP Q13469
M	629	ALA	ASP	CONFLICT	UNP Q13469
M	630	ALA	LYS	CONFLICT	UNP Q13469
M	631	ALA	ASP	CONFLICT	UNP Q13469
M	632	ALA	LYS	CONFLICT	UNP Q13469
M	633	ALA	SER	CONFLICT	UNP Q13469
M	634	ALA	GLN	CONFLICT	UNP Q13469

- Molecule 2 is a protein called Forkhead box protein P3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	F	82	Total	C	N	O	S	0	0	0
			714	465	132	114	3			

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	G	82	Total	C	N	O	S	0	0	0
			714	465	132	114	3			
2	H	82	Total	C	N	O	S	0	0	0
			714	465	132	114	3			
2	I	82	Total	C	N	O	S	0	0	0
			714	465	132	114	3			

- Molecule 3 is a DNA chain called human hARRE2 DNA (Plus Strand).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	21	Total	C	N	O	P	0	0	0
			431	209	76	126	20			
3	A	21	Total	C	N	O	P	0	0	0
			431	209	76	126	20			

- Molecule 4 is a DNA chain called human hARRE2 DNA (Minus Strand).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	21	Total	C	N	O	P	0	0	0
			424	206	76	122	20			
4	B	21	Total	C	N	O	P	0	0	0
			424	206	76	122	20			

- Molecule 5 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	H	1	Total	Mg	0	0
			1	1		
5	G	1	Total	Mg	0	0
			1	1		
5	I	1	Total	Mg	0	0
			1	1		
5	F	1	Total	Mg	0	0
			1	1		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	N	1	Total	O	0	0
			1	1		
6	C	2	Total	O	0	0
			2	2		

*Continued on next page...*

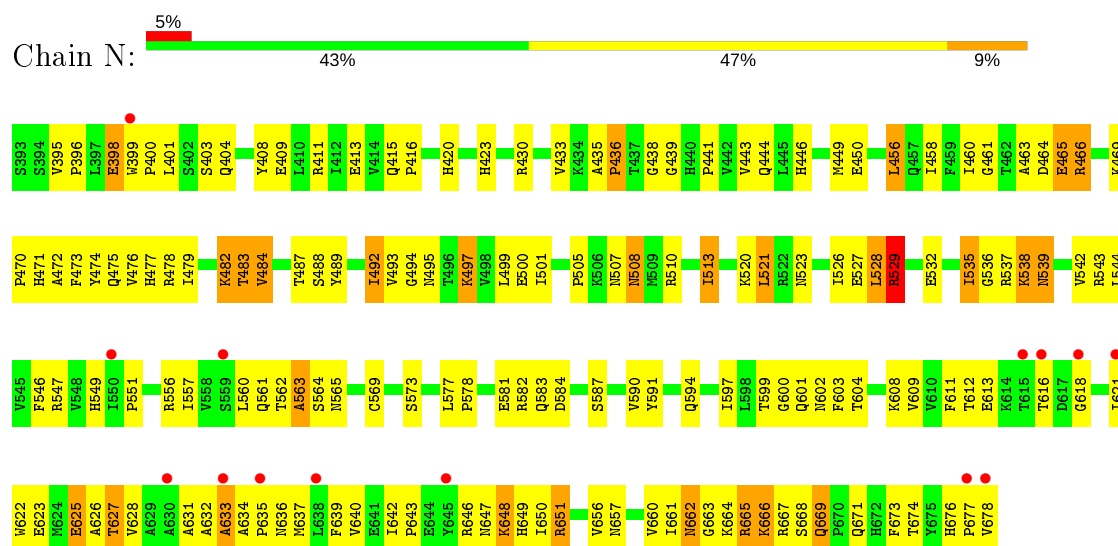
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	D	3	Total	O	0	0
			3	3		
6	B	1	Total	O	0	0
			1	1		

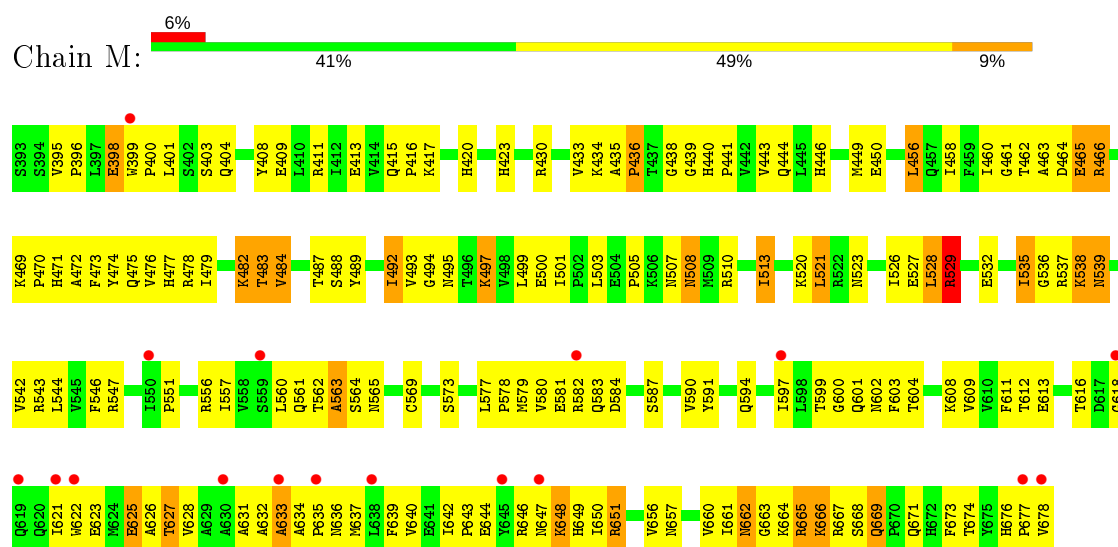
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA 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: Nuclear factor of activated T-cells, cytoplasmic 2

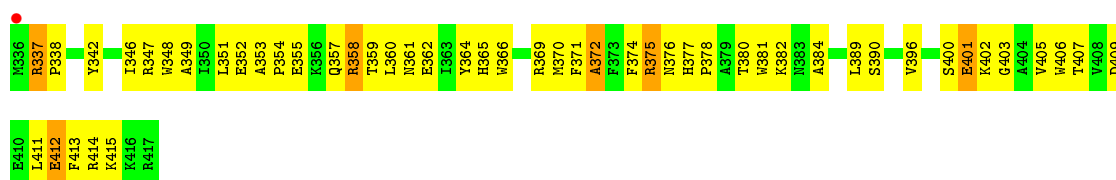


- Molecule 1: Nuclear factor of activated T-cells, cytoplasmic 2

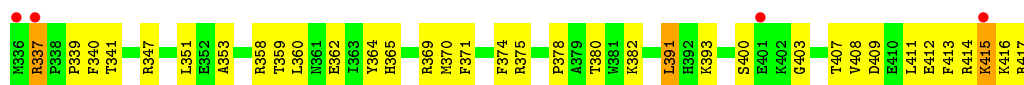


- Molecule 2: Forkhead box protein P3

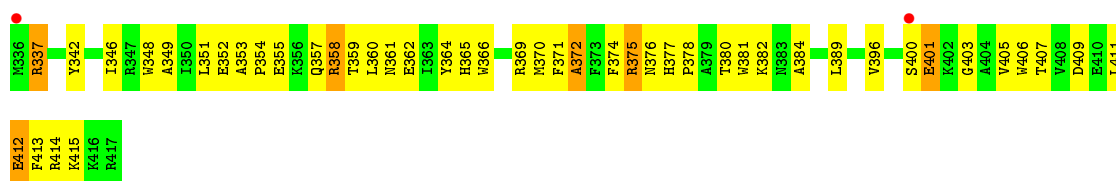
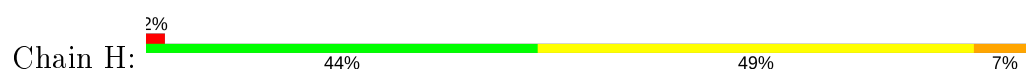




- Molecule 2: Forkhead box protein P3



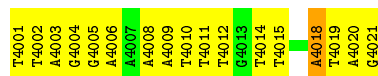
- Molecule 2: Forkhead box protein P3



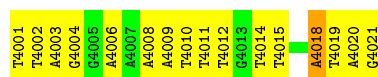
- Molecule 2: Forkhead box protein P3



- Molecule 3: human hARRE2 DNA (Plus Strand)



- Molecule 3: human hARRE2 DNA (Plus Strand)

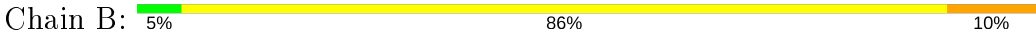


- Molecule 4: human hARRE2 DNA (Minus Strand)



A5001	A5002	C5003	T5004	A5005	T5006	G5007	A5008	A5009	A5010	C5011	A5012	A5013	A5014	T5015	T5016	T5017	T5018	C5019	C5020	T5021
-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

● Molecule 4: human hARRE2 DNA (Minus Strand)



A5001	A5002	C5003	T5004	A5005	T5006	G5007	A5008	A5009	A5010	C5011	A5012	A5013	A5014	T5015	T5016	T5017	T5018	C5019	C5020	T5021
-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	83.67Å 131.23Å 68.67Å 90.00° 89.95° 90.00°	Depositor
Resolution (Å)	50.00 – 2.80 19.93 – 2.80	Depositor EDS
% Data completeness (in resolution range)	(Not available) (50.00-2.80) 96.3 (19.93-2.80)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.68 (at 2.79Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.248 , 0.283 0.244 , 0.278	Depositor DCC
$R_{free}$ test set	3538 reflections (9.74%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	63.7	Xtriage
Anisotropy	0.645	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , 23.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.480 for -h,-k,l	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	9105	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	65.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.89% 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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: MG

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	M	0.38	0/2314	0.65	0/3134
1	N	0.39	0/2314	0.65	0/3134
2	F	0.49	0/738	0.63	0/996
2	G	0.48	0/738	0.62	0/996
2	H	0.50	0/738	0.63	0/996
2	I	0.47	0/738	0.62	0/996
3	A	0.58	0/483	0.89	1/745 (0.1%)
3	C	0.53	0/483	0.88	1/745 (0.1%)
4	B	0.57	0/475	0.82	0/730
4	D	0.55	0/475	0.81	0/730
All	All	0.46	0/9496	0.69	2/13202 (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
3	A	0	1
3	C	0	1
4	B	0	2
4	D	0	2
All	All	0	6

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	4006	DA	N9-C1'-C2'	5.36	122.78	112.60
3	C	4006	DA	N9-C1'-C2'	5.25	122.56	112.60

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	A	4018	DA	Sidechain
4	B	5004	DT	Sidechain
3	C	4018	DA	Sidechain
4	D	5004	DT	Sidechain
4	D	5005	DA	Sidechain

## 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	M	2264	0	2271	168	0
1	N	2264	0	2271	161	0
2	F	714	0	708	51	0
2	G	714	0	708	38	0
2	H	714	0	708	48	0
2	I	714	0	708	38	0
3	A	431	0	242	24	0
3	C	431	0	242	25	0
4	B	424	0	240	24	0
4	D	424	0	240	25	0
5	F	1	0	0	0	0
5	G	1	0	0	0	0
5	H	1	0	0	0	0
5	I	1	0	0	0	0
6	B	1	0	0	0	0
6	C	2	0	0	0	0
6	D	3	0	0	0	0
6	N	1	0	0	0	0
All	All	9105	0	8338	563	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 32.

The worst 5 of 563 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:M:634:ALA:HB3	1:M:637:MET:HB2	1.44	0.98
1:N:669:GLN:HE21	1:N:671:GLN:HE22	1.07	0.98
3:A:4002:DT:H2"	3:A:4003:DA:H5"	1.44	0.98
1:N:475:GLN:HE22	1:N:497:LYS:H	1.04	0.98
3:C:4002:DT:H2"	3:C:4003:DA:H5"	1.45	0.97

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	M	284/286 (99%)	216 (76%)	42 (15%)	26 (9%)	1	1
1	N	284/286 (99%)	217 (76%)	41 (14%)	26 (9%)	1	1
2	F	80/82 (98%)	66 (82%)	10 (12%)	4 (5%)	2	6
2	G	80/82 (98%)	69 (86%)	11 (14%)	0	100	100
2	H	80/82 (98%)	66 (82%)	10 (12%)	4 (5%)	2	6
2	I	80/82 (98%)	69 (86%)	11 (14%)	0	100	100
All	All	888/900 (99%)	703 (79%)	125 (14%)	60 (7%)	1	3

5 of 60 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	N	465	GLU
1	N	563	ALA
1	N	583	GLN
1	N	604	THR
1	N	627	THR

### 5.3.2 Protein sidechains ⓘ

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	M	250/250 (100%)	230 (92%)	20 (8%)	12	34
1	N	250/250 (100%)	230 (92%)	20 (8%)	12	34
2	F	74/74 (100%)	72 (97%)	2 (3%)	44	78
2	G	74/74 (100%)	69 (93%)	5 (7%)	16	42
2	H	74/74 (100%)	72 (97%)	2 (3%)	44	78
2	I	74/74 (100%)	69 (93%)	5 (7%)	16	42
All	All	796/796 (100%)	742 (93%)	54 (7%)	16	42

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

Mol	Chain	Res	Type
2	G	380	THR
1	M	436	PRO
2	I	337	ARG
2	G	391	LEU
1	M	395	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 33 such sidechains are listed below:

Mol	Chain	Res	Type
2	G	365	HIS
1	M	457	GLN
2	I	357	GLN
2	G	377	HIS
1	M	423	HIS

### 5.3.3 RNA ⓘ

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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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<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	M	286/286 (100%)	0.20	17 (5%) 22 14	36, 78, 110, 117	0
1	N	286/286 (100%)	0.19	14 (4%) 29 20	36, 78, 110, 117	0
2	F	82/82 (100%)	-0.01	1 (1%) 79 73	31, 50, 82, 98	0
2	G	82/82 (100%)	0.10	4 (4%) 29 20	28, 52, 95, 100	0
2	H	82/82 (100%)	0.06	2 (2%) 59 49	31, 50, 81, 98	0
2	I	82/82 (100%)	0.10	2 (2%) 59 49	29, 52, 95, 100	0
3	A	21/21 (100%)	-0.56	0 100 100	32, 42, 65, 71	0
3	C	21/21 (100%)	-0.50	0 100 100	31, 42, 65, 71	0
4	B	21/21 (100%)	-0.35	0 100 100	32, 43, 68, 81	0
4	D	21/21 (100%)	-0.40	0 100 100	32, 43, 68, 81	0
All	All	984/984 (100%)	0.10	40 (4%) 37 27	28, 66, 105, 117	0

The worst 5 of 40 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	336	MET	5.0
2	H	336	MET	4.2
1	N	618	GLY	4.1
2	I	336	MET	3.7
2	I	415	LYS	3.4

### 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 [i](#)

There are no carbohydrates in this entry.

### 6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
5	MG	I	1	1/1	0.81	0.22	61,61,61,61	0
5	MG	F	1	1/1	0.82	0.09	43,43,43,43	0
5	MG	G	1	1/1	0.85	0.19	62,62,62,62	0
5	MG	H	1	1/1	0.86	0.09	43,43,43,43	0

### 6.5 Other polymers [i](#)

There are no such residues in this entry.