



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 15, 2020 – 08:58 pm BST

PDB ID : 3O4X  
Title : Crystal structure of complex between amino and carboxy terminal fragments of mDia1  
Authors : Eck, M.J.; Nezami, A.; Toms, A.V.  
Deposited on : 2010-07-27  
Resolution : 3.20 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

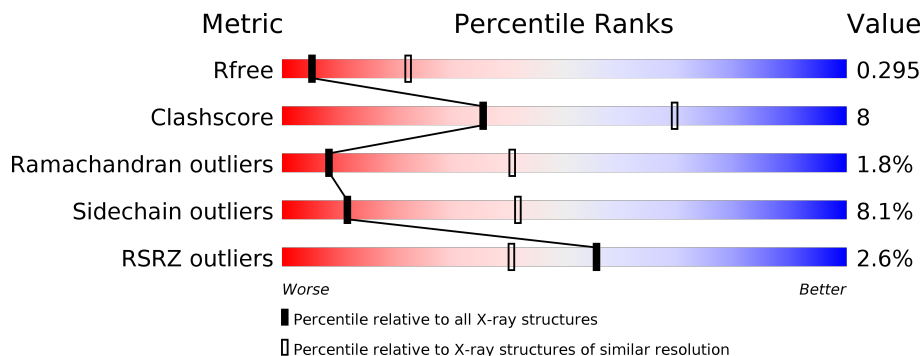
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.20 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	1133 (3.20-3.20)
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)
RSRZ outliers	127900	1095 (3.20-3.20)

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	A	330	
1	B	330	
1	C	330	
1	D	330	
2	E	467	
2	F	467	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	G	467	 <p>3% 70% 19% • 7%</p>
2	H	467	 <p>2% 72% 16% • 7%</p>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 24335 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 Protein diaphanous homolog 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	317	2552	1605	440	485	22	0	0	0
1	B	318	2559	1609	441	487	22	0	0	0
1	C	311	2508	1579	432	475	22	0	0	0
1	D	319	2568	1614	442	490	22	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	129	GLY	-	EXPRESSION TAG	UNP O08808
A	130	SER	-	EXPRESSION TAG	UNP O08808
B	129	GLY	-	EXPRESSION TAG	UNP O08808
B	130	SER	-	EXPRESSION TAG	UNP O08808
C	129	GLY	-	EXPRESSION TAG	UNP O08808
C	130	SER	-	EXPRESSION TAG	UNP O08808
D	129	GLY	-	EXPRESSION TAG	UNP O08808
D	130	SER	-	EXPRESSION TAG	UNP O08808

- Molecule 2 is a protein called Protein diaphanous homolog 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	E	433	3525	2230	602	672	21	0	0	0
2	H	433	3525	2230	602	672	21	0	0	0
2	G	433	3525	2230	602	672	21	0	0	0
2	F	433	3525	2230	602	672	21	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	734	GLY	-	EXPRESSION TAG	UNP O08808
E	735	SER	-	EXPRESSION TAG	UNP O08808
H	734	GLY	-	EXPRESSION TAG	UNP O08808
H	735	SER	-	EXPRESSION TAG	UNP O08808
G	734	GLY	-	EXPRESSION TAG	UNP O08808
G	735	SER	-	EXPRESSION TAG	UNP O08808
F	734	GLY	-	EXPRESSION TAG	UNP O08808
F	735	SER	-	EXPRESSION TAG	UNP O08808

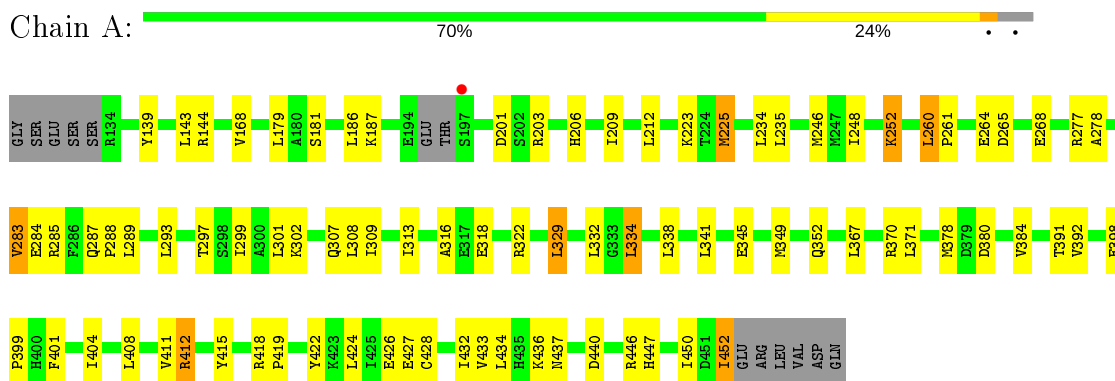
- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	8	Total O 8 8	0	0
3	B	9	Total O 9 9	0	0
3	C	6	Total O 6 6	0	0
3	D	8	Total O 8 8	0	0
3	E	5	Total O 5 5	0	0
3	H	4	Total O 4 4	0	0
3	G	6	Total O 6 6	0	0
3	F	2	Total O 2 2	0	0

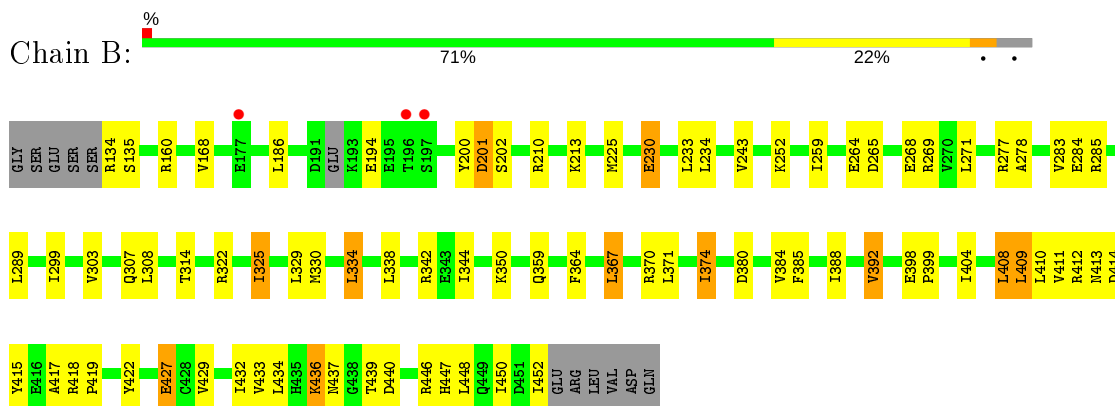
### 3 Residue-property plots [i](#)

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.

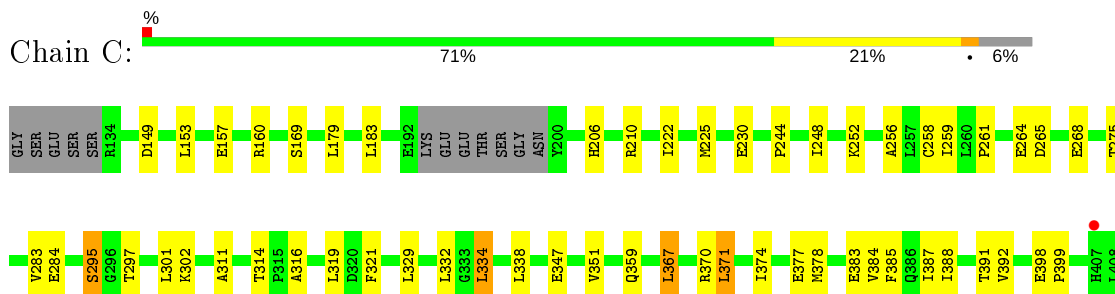
- Molecule 1: Protein diaphanous homolog 1

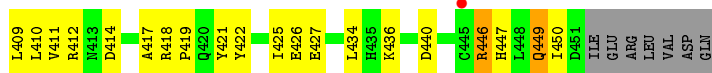


- Molecule 1: Protein diaphanous homolog 1

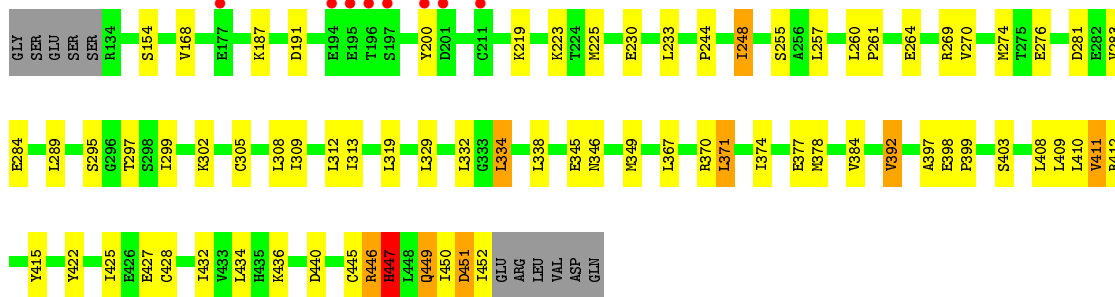
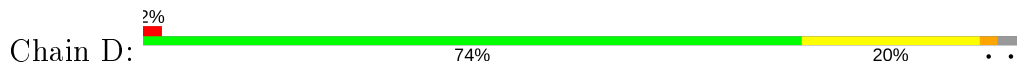


- Molecule 1: Protein diaphanous homolog 1

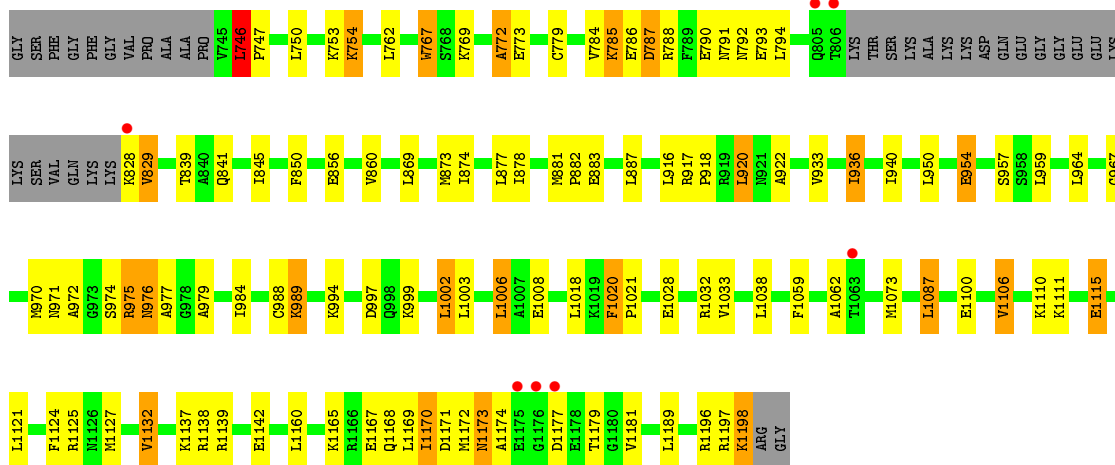




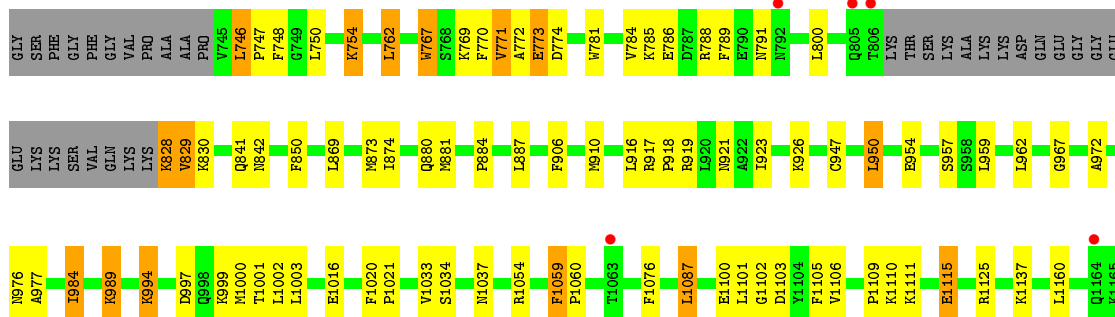
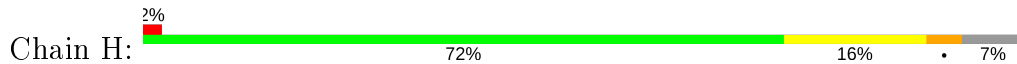
• Molecule 1: Protein diaphanous homolog 1

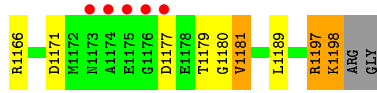


• Molecule 2: Protein diaphanous homolog 1

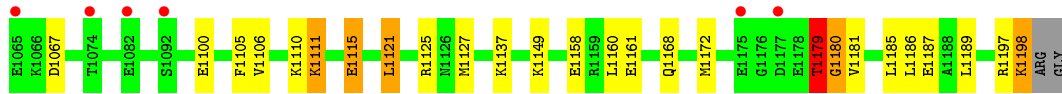


• Molecule 2: Protein diaphanous homolog 1

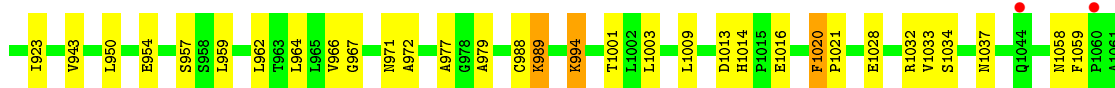
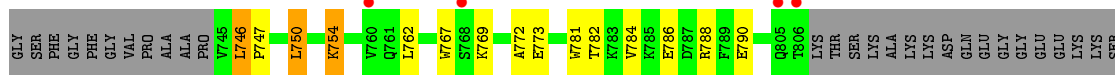




• Molecule 2: Protein diaphanous homolog 1



• Molecule 2: Protein diaphanous homolog 1





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	95.76Å 206.83Å 131.06Å 90.00° 105.94° 90.00°	Depositor
Resolution (Å)	20.12 – 3.20 20.12 – 3.20	Depositor EDS
% Data completeness (in resolution range)	100.0 (20.12-3.20) 99.1 (20.12-3.20)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.11	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.45 (at 3.22Å)	Xtrriage
Refinement program	REFMAC 5.5.0109	Depositor
R, $R_{free}$	0.233 , 0.298 0.232 , 0.295	Depositor DCC
$R_{free}$ test set	3981 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	66.5	Xtrriage
Anisotropy	0.118	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 77.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.44$ , $\langle L^2 \rangle = 0.27$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	24335	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	107.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.73% 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	0.36	0/2588	0.58	0/3483
1	B	0.37	0/2595	0.56	0/3493
1	C	0.37	0/2544	0.55	1/3425 (0.0%)
1	D	0.36	0/2605	0.55	1/3508 (0.0%)
2	E	0.37	0/3580	0.55	0/4799
2	F	0.36	0/3580	0.53	0/4799
2	G	0.36	0/3580	0.53	0/4799
2	H	0.38	0/3580	0.53	0/4799
All	All	0.37	0/24652	0.55	2/33105 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	371	LEU	CA-CB-CG	6.41	130.03	115.30
1	D	371	LEU	CA-CB-CG	5.18	127.21	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	2552	0	2563	59	0
1	B	2559	0	2570	73	0
1	C	2508	0	2519	57	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	D	2568	0	2577	47	0
2	E	3525	0	3552	65	0
2	F	3525	0	3552	57	0
2	G	3525	0	3552	67	0
2	H	3525	0	3552	52	0
3	A	8	0	0	0	0
3	B	9	0	0	1	0
3	C	6	0	0	0	0
3	D	8	0	0	0	0
3	E	5	0	0	0	0
3	F	2	0	0	0	0
3	G	6	0	0	1	0
3	H	4	0	0	0	0
All	All	24335	0	24437	396	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 8.

The worst 5 of 396 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)
2:E:988:CYS:HA	2:E:1127:MET:HE1	1.34	1.09
1:D:446:ARG:O	1:D:447:HIS:HB2	1.66	0.94
2:G:754:LYS:H	2:G:754:LYS:HE2	1.32	0.94
1:B:410:LEU:HD11	1:C:371:LEU:HB2	1.47	0.94
2:E:828:LYS:N	2:E:829:VAL:HA	1.81	0.92

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	313/330 (95%)	299 (96%)	13 (4%)	1 (0%)	41	74
1	B	314/330 (95%)	296 (94%)	15 (5%)	3 (1%)	15	54
1	C	307/330 (93%)	289 (94%)	14 (5%)	4 (1%)	12	47
1	D	317/330 (96%)	296 (93%)	16 (5%)	5 (2%)	9	43
2	E	429/467 (92%)	390 (91%)	26 (6%)	13 (3%)	4	28
2	F	429/467 (92%)	387 (90%)	33 (8%)	9 (2%)	7	37
2	G	429/467 (92%)	389 (91%)	31 (7%)	9 (2%)	7	37
2	H	429/467 (92%)	384 (90%)	37 (9%)	8 (2%)	8	39
All	All	2967/3188 (93%)	2730 (92%)	185 (6%)	52 (2%)	8	41

5 of 52 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	264	GLU
1	C	264	GLU
1	D	447	HIS
1	D	451	ASP
2	E	975	ARG

### 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	287/299 (96%)	264 (92%)	23 (8%)	12	42
1	B	288/299 (96%)	270 (94%)	18 (6%)	18	52
1	C	282/299 (94%)	264 (94%)	18 (6%)	17	52
1	D	289/299 (97%)	262 (91%)	27 (9%)	9	33
2	E	391/416 (94%)	351 (90%)	40 (10%)	7	29
2	F	391/416 (94%)	363 (93%)	28 (7%)	14	47
2	G	391/416 (94%)	358 (92%)	33 (8%)	11	39
2	H	391/416 (94%)	358 (92%)	33 (8%)	11	39

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	2710/2860 (95%)	2490 (92%)	220 (8%)	11 42

5 of 220 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	E	936	ILE
2	E	1170	ILE
2	F	989	LYS
2	E	959	LEU
2	E	1059	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 43 such sidechains are listed below:

Mol	Chain	Res	Type
2	E	1168	GLN
2	H	976	ASN
2	F	932	GLN
2	H	791	ASN
2	H	841	GLN

### 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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

There are no ligands in this entry.

## 5.7 Other polymers

There are no such residues in this entry.

## 5.8 Polymer linkage issues

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	317/330 (96%)	-0.45	1 (0%) 94 92	30, 75, 117, 149	2 (0%)
1	B	318/330 (96%)	-0.45	3 (0%) 84 75	38, 79, 131, 157	2 (0%)
1	C	311/330 (94%)	-0.45	2 (0%) 89 83	30, 80, 129, 213	2 (0%)
1	D	319/330 (96%)	-0.14	8 (2%) 57 43	31, 112, 166, 198	2 (0%)
2	E	433/467 (92%)	-0.23	7 (1%) 72 59	47, 94, 144, 177	0
2	F	433/467 (92%)	0.22	32 (7%) 14 8	61, 133, 211, 256	0
2	G	433/467 (92%)	0.08	14 (3%) 47 31	68, 127, 172, 209	0
2	H	433/467 (92%)	-0.08	10 (2%) 60 47	40, 110, 166, 200	0
All	All	2997/3188 (94%)	-0.16	77 (2%) 56 40	30, 102, 173, 256	8 (0%)

The worst 5 of 77 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	899	ASP	7.9
2	F	806	THR	6.5
2	F	831	GLU	6.4
2	F	832	LEU	4.8
2	H	1176	GLY	4.3

### 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 carbohydrates 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.