



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 09:30 AM GMT

PDB ID : 3IOT  
Title : Huntingtin amino-terminal region with 17 Gln residues - crystal C92-b  
Authors : Kim, M.W.  
Deposited on : 2009-08-14  
Resolution : 3.50 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  
<http://wwpdb.org/validation/2016/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  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk26865

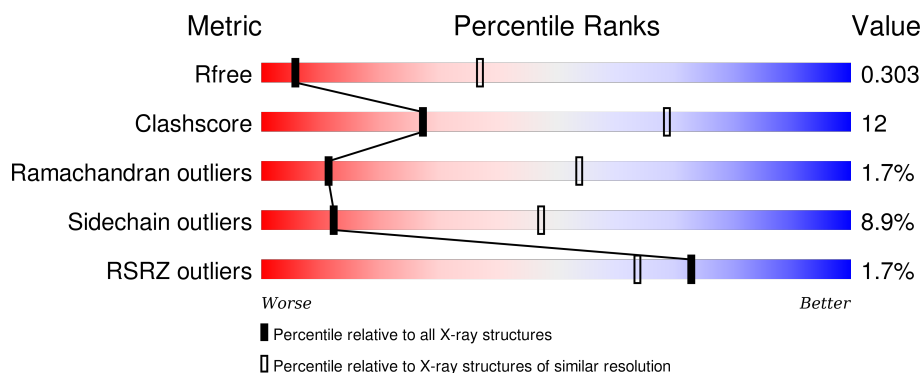
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.50 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))
$R_{free}$	91344	1051 (3.60-3.40)
Clashscore	102246	1157 (3.60-3.40)
Ramachandran outliers	100387	1120 (3.60-3.40)
Sidechain outliers	100360	1121 (3.60-3.40)
RSRZ outliers	91569	1058 (3.60-3.40)

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. 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	449	<div> <div>2%</div> <div>63% 20% • 14%</div> </div>
1	B	449	<div> <div>2%</div> <div>67% 19% • 11%</div> </div>
1	C	449	<div> <div>2%</div> <div>67% 22% • • 7%</div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 9195 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 Maltose-binding protein, huntingtin fusion protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	388	Total	C	N	O	S	33	0	0
			2967	1911	482	566	8			
1	B	399	Total	C	N	O	S	12	0	0
			3044	1952	497	587	8			
1	C	418	Total	C	N	O	S	18	0	0
			3166	2038	516	604	8			

There are 108 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	360	ALA	-	LINKER	? ?
A	361	LEU	-	LINKER	? ?
A	362	ALA	-	LINKER	? ?
A	363	ALA	-	LINKER	? ?
A	364	ALA	-	LINKER	? ?
A	365	GLN	-	LINKER	? ?
A	366	THR	-	LINKER	? ?
A	367	ASN	-	LINKER	? ?
A	368	ALA	-	LINKER	? ?
A	369	ALA	-	LINKER	? ?
A	370	ALA	-	LINKER	? ?
A	?	-	GLN	DELETION	UNP P42858
A	?	-	GLN	DELETION	UNP P42858
A	?	-	GLN	DELETION	UNP P42858
A	?	-	GLN	DELETION	UNP P42858
A	?	-	GLN	DELETION	UNP P42858
A	?	-	GLN	DELETION	UNP P42858
A	431	GLN	-	EXPRESSION TAG	? ?
A	432	SER	-	EXPRESSION TAG	? ?
A	433	TYR	-	EXPRESSION TAG	? ?
A	434	GLN	-	EXPRESSION TAG	? ?
A	435	ILE	-	EXPRESSION TAG	? ?
A	436	THR	-	EXPRESSION TAG	? ?

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	437	ALA	-	EXPRESSION TAG	? ?
A	438	GLY	-	EXPRESSION TAG	? ?
A	439	LYS	-	EXPRESSION TAG	? ?
A	440	LEU	-	EXPRESSION TAG	? ?
A	441	GLY	-	EXPRESSION TAG	? ?
A	442	THR	-	EXPRESSION TAG	? ?
A	443	GLY	-	EXPRESSION TAG	? ?
A	444	ARG	-	EXPRESSION TAG	? ?
A	445	ARG	-	EXPRESSION TAG	? ?
A	446	PHE	-	EXPRESSION TAG	? ?
A	447	THR	-	EXPRESSION TAG	? ?
A	448	THR	-	EXPRESSION TAG	? ?
A	449	SER	-	EXPRESSION TAG	? ?
B	360	ALA	-	LINKER	? ?
B	361	LEU	-	LINKER	? ?
B	362	ALA	-	LINKER	? ?
B	363	ALA	-	LINKER	? ?
B	364	ALA	-	LINKER	? ?
B	365	GLN	-	LINKER	? ?
B	366	THR	-	LINKER	? ?
B	367	ASN	-	LINKER	? ?
B	368	ALA	-	LINKER	? ?
B	369	ALA	-	LINKER	? ?
B	370	ALA	-	LINKER	? ?
B	?	-	GLN	DELETION	UNP P42858
B	?	-	GLN	DELETION	UNP P42858
B	?	-	GLN	DELETION	UNP P42858
B	?	-	GLN	DELETION	UNP P42858
B	?	-	GLN	DELETION	UNP P42858
B	?	-	GLN	DELETION	UNP P42858
B	431	GLN	-	EXPRESSION TAG	? ?
B	432	SER	-	EXPRESSION TAG	? ?
B	433	TYR	-	EXPRESSION TAG	? ?
B	434	GLN	-	EXPRESSION TAG	? ?
B	435	ILE	-	EXPRESSION TAG	? ?
B	436	THR	-	EXPRESSION TAG	? ?
B	437	ALA	-	EXPRESSION TAG	? ?
B	438	GLY	-	EXPRESSION TAG	? ?
B	439	LYS	-	EXPRESSION TAG	? ?
B	440	LEU	-	EXPRESSION TAG	? ?
B	441	GLY	-	EXPRESSION TAG	? ?
B	442	THR	-	EXPRESSION TAG	? ?

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	443	GLY	-	EXPRESSION TAG	? ?
B	444	ARG	-	EXPRESSION TAG	? ?
B	445	ARG	-	EXPRESSION TAG	? ?
B	446	PHE	-	EXPRESSION TAG	? ?
B	447	THR	-	EXPRESSION TAG	? ?
B	448	THR	-	EXPRESSION TAG	? ?
B	449	SER	-	EXPRESSION TAG	? ?
C	360	ALA	-	LINKER	? ?
C	361	LEU	-	LINKER	? ?
C	362	ALA	-	LINKER	? ?
C	363	ALA	-	LINKER	? ?
C	364	ALA	-	LINKER	? ?
C	365	GLN	-	LINKER	? ?
C	366	THR	-	LINKER	? ?
C	367	ASN	-	LINKER	? ?
C	368	ALA	-	LINKER	? ?
C	369	ALA	-	LINKER	? ?
C	370	ALA	-	LINKER	? ?
C	?	-	GLN	DELETION	UNP P42858
C	?	-	GLN	DELETION	UNP P42858
C	?	-	GLN	DELETION	UNP P42858
C	?	-	GLN	DELETION	UNP P42858
C	?	-	GLN	DELETION	UNP P42858
C	?	-	GLN	DELETION	UNP P42858
C	431	GLN	-	EXPRESSION TAG	? ?
C	432	SER	-	EXPRESSION TAG	? ?
C	433	TYR	-	EXPRESSION TAG	? ?
C	434	GLN	-	EXPRESSION TAG	? ?
C	435	ILE	-	EXPRESSION TAG	? ?
C	436	THR	-	EXPRESSION TAG	? ?
C	437	ALA	-	EXPRESSION TAG	? ?
C	438	GLY	-	EXPRESSION TAG	? ?
C	439	LYS	-	EXPRESSION TAG	? ?
C	440	LEU	-	EXPRESSION TAG	? ?
C	441	GLY	-	EXPRESSION TAG	? ?
C	442	THR	-	EXPRESSION TAG	? ?
C	443	GLY	-	EXPRESSION TAG	? ?
C	444	ARG	-	EXPRESSION TAG	? ?
C	445	ARG	-	EXPRESSION TAG	? ?
C	446	PHE	-	EXPRESSION TAG	? ?
C	447	THR	-	EXPRESSION TAG	? ?
C	448	THR	-	EXPRESSION TAG	? ?

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	449	SER	-	EXPRESSION TAG	? ?

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	3	Total 3	Zn 3	0	0
2	A	3	Total 3	Zn 3	0	0
2	C	3	Total 3	Zn 3	0	0

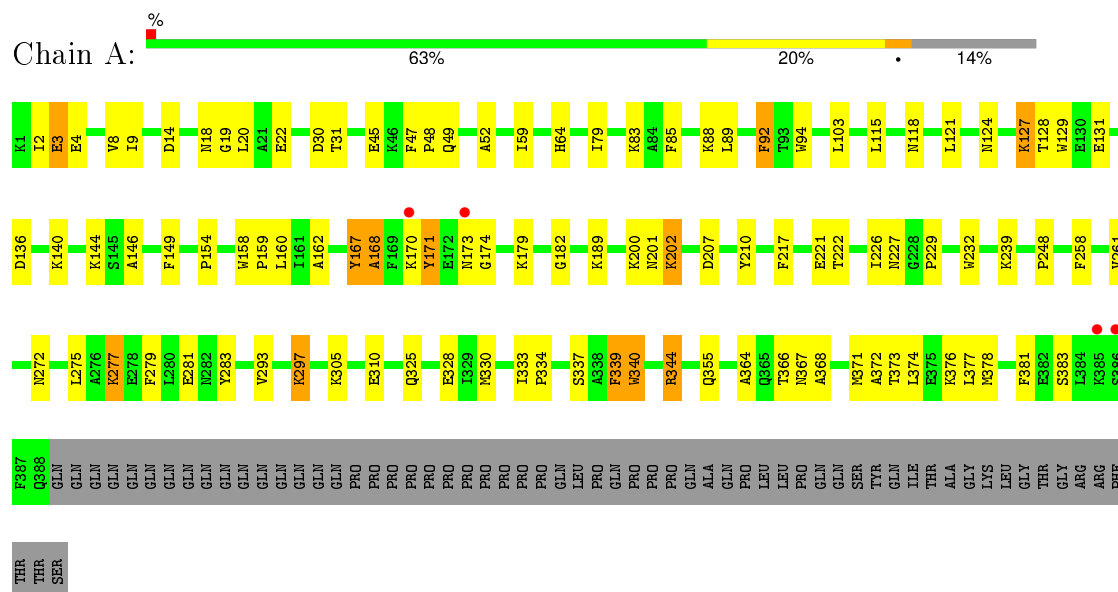
- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	2	Total 2	Ca 2	0	0
3	A	5	Total 5	Ca 5	0	0
3	C	2	Total 2	Ca 2	0	0

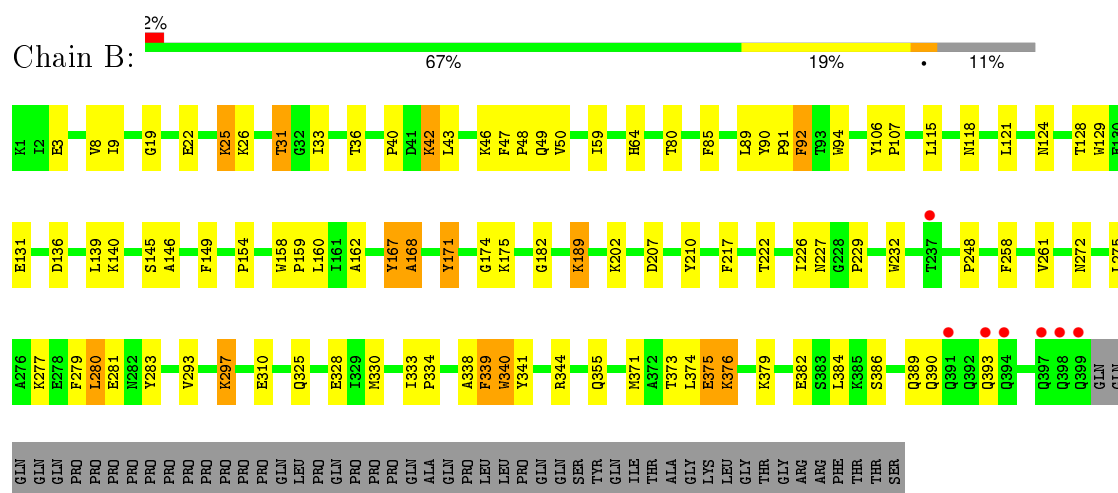
### 3 Residue-property plots

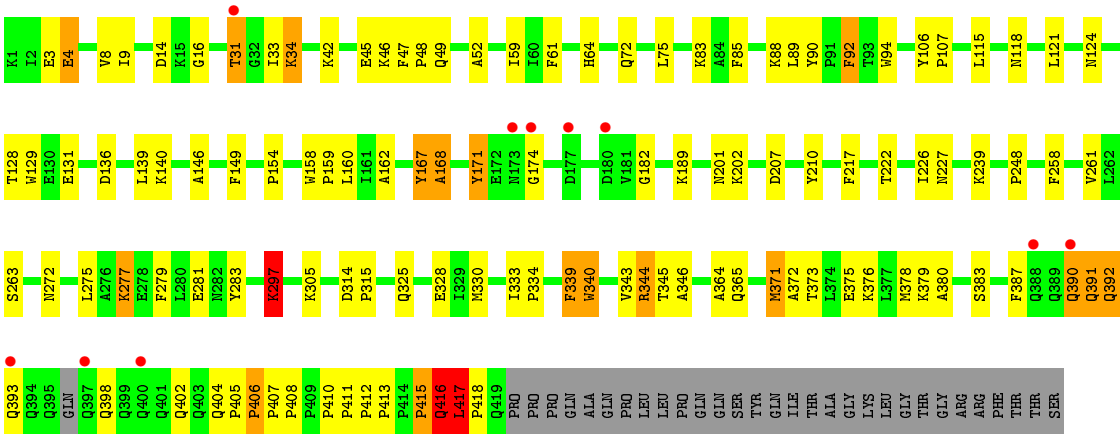
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according 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: Maltose-binding protein, huntingtin fusion protein



- Molecule 1: Maltose-binding protein, huntingtin fusion protein







## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	163.25Å 101.36Å 137.03Å 90.00° 91.83° 90.00°	Depositor
Resolution (Å)	37.36 – 3.50 29.87 – 3.50	Depositor EDS
% Data completeness (in resolution range)	69.2 (37.36-3.50) 69.3 (29.87-3.50)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.09	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.68 (at 3.47Å)	Xtriage
Refinement program	REFMAC 5.5.0102	Depositor
R, $R_{free}$	0.247 , 0.295 0.258 , 0.303	Depositor DCC
$R_{free}$ test set	975 reflections (5.22%)	DCC
Wilson B-factor (Å <sup>2</sup> )	79.8	Xtriage
Anisotropy	0.222	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.24 , -10.0	EDS
Estimated twinning fraction	0.024 for -h,-k,l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	1 of 19645 reflections (0.005%)	Xtriage
$F_o, F_c$ correlation	0.88	EDS
Total number of atoms	9195	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	58.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.18% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: ZN, CA

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	0.61	7/3038 (0.2%)	0.65	7/4130 (0.2%)
1	B	0.53	4/3115 (0.1%)	0.62	6/4238 (0.1%)
1	C	0.53	2/3248 (0.1%)	0.71	9/4430 (0.2%)
All	All	0.56	13/9401 (0.1%)	0.66	22/12798 (0.2%)

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	C	0	1

The worst 5 of 13 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	179	LYS	CB-CG	14.45	1.91	1.52
1	C	392	GLN	C-O	-13.95	0.96	1.23
1	B	26	LYS	CB-CG	11.75	1.84	1.52
1	A	378	MET	CB-CG	9.58	1.81	1.51
1	A	88	LYS	CB-CG	-8.86	1.28	1.52

The worst 5 of 22 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	391	GLN	O-C-N	-18.06	93.80	122.70
1	C	46	LYS	CA-CB-CG	-14.82	80.79	113.40
1	A	378	MET	CA-CB-CG	-12.49	92.07	113.30
1	B	297	LYS	CD-CE-NZ	-11.26	85.80	111.70
1	A	297	LYS	CD-CE-NZ	-10.15	88.36	111.70

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	391	GLN	Mainchain

## 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	A	2967	0	2908	64	0
1	B	3044	0	2946	65	0
1	C	3166	0	3062	90	0
2	A	3	0	0	0	1
2	B	3	0	0	0	1
2	C	3	0	0	0	0
3	A	5	0	0	0	1
3	B	2	0	0	0	0
3	C	2	0	0	0	0
All	All	9195	0	8916	213	2

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 12.

The worst 5 of 213 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:C:404:GLN:C	1:C:406:PRO:HD3	1.53	1.28
1:C:404:GLN:O	1:C:406:PRO:HD3	1.45	1.16
1:C:404:GLN:O	1:C:406:PRO:CD	1.94	1.15
1:C:416:GLN:HA	1:C:417:LEU:CB	1.83	1.09
1:C:416:GLN:HA	1:C:417:LEU:HB3	1.49	0.95

All (2) 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)
2:A:502:ZN:ZN	2:B:503:ZN:ZN[3_455]	0.99	1.21
3:A:504:CA:CA	3:A:508:CA:CA[2_455]	1.40	0.80

## 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	386/449 (86%)	366 (95%)	16 (4%)	4 (1%)	19 66
1	B	397/449 (88%)	374 (94%)	20 (5%)	3 (1%)	24 70
1	C	414/449 (92%)	380 (92%)	21 (5%)	13 (3%)	5 41
All	All	1197/1347 (89%)	1120 (94%)	57 (5%)	20 (2%)	11 54

5 of 20 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	3	GLU
1	B	3	GLU
1	C	3	GLU
1	C	415	PRO
1	C	417	LEU

### 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	298/363 (82%)	270 (91%)	28 (9%)	11 44

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	304/363 (84%)	280 (92%)	24 (8%)	15	53
1	C	317/363 (87%)	287 (90%)	30 (10%)	11	43
All	All	919/1089 (84%)	837 (91%)	82 (9%)	12	47

5 of 82 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	124	ASN
1	B	339	PHE
1	C	371	MET
1	B	160	LEU
1	B	189	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	201	ASN
1	B	218	ASN
1	C	64	HIS
1	B	64	HIS
1	C	72	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](#)

Of 18 ligands modelled in this entry, 18 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 [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	388/449 (86%)	-0.42	4 (1%) 84 76	55, 55, 63, 88	24 (6%)
1	B	399/449 (88%)	-0.28	7 (1%) 71 62	55, 55, 87, 87	37 (9%)
1	C	418/449 (93%)	-0.33	10 (2%) 62 52	55, 55, 87, 88	43 (10%)
All	All	1205/1347 (89%)	-0.34	21 (1%) 73 64	55, 55, 87, 88	104 (8%)

The worst 5 of 21 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	397	GLN	4.1
1	B	397	GLN	3.4
1	B	399	GLN	3.3
1	C	388	GLN	3.0
1	B	391	GLN	2.8

### 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 [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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	ZN	A	501	1/1	0.98	0.16	-	76,76,76,76	0
3	CA	A	508	1/1	0.97	0.14	-	80,80,80,80	0
3	CA	B	504	1/1	0.86	0.32	-	70,70,70,70	0
3	CA	A	506	1/1	0.97	0.28	-	62,62,62,62	0
2	ZN	A	503	1/1	0.94	0.12	-	98,98,98,98	0
2	ZN	B	501	1/1	0.94	0.20	-	98,98,98,98	0
3	CA	C	505	1/1	0.89	0.29	-	75,75,75,75	0
2	ZN	C	503	1/1	0.88	0.07	-	167,167,167,167	0
3	CA	B	505	1/1	0.96	0.37	-	86,86,86,86	0
2	ZN	C	502	1/1	0.87	0.12	-	103,103,103,103	0
3	CA	A	505	1/1	0.98	0.28	-	32,32,32,32	0
2	ZN	A	502	1/1	0.95	0.12	-	99,99,99,99	0
3	CA	A	504	1/1	0.99	0.16	-	69,69,69,69	0
2	ZN	B	502	1/1	0.98	0.28	-	78,78,78,78	0
3	CA	C	504	1/1	0.88	0.42	-	96,96,96,96	0
2	ZN	C	501	1/1	0.84	0.11	-	132,132,132,132	0
2	ZN	B	503	1/1	0.95	0.12	-	110,110,110,110	0
3	CA	A	507	1/1	0.85	0.19	-	87,87,87,87	0

## 6.5 Other polymers

There are no such residues in this entry.