



# wwPDB X-ray Structure Validation Summary Report i

Feb 26, 2014 – 04:43 PM GMT

PDB ID : 3IOV  
Title : Huntingtin amino-terminal region with 17 Gln residues - crystal C99  
Authors : Kim, M.W.  
Deposited on : 2009-08-14  
Resolution : 3.70 Å(reported)

This is a wwPDB 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/ValidationPDFNotes.html>

---

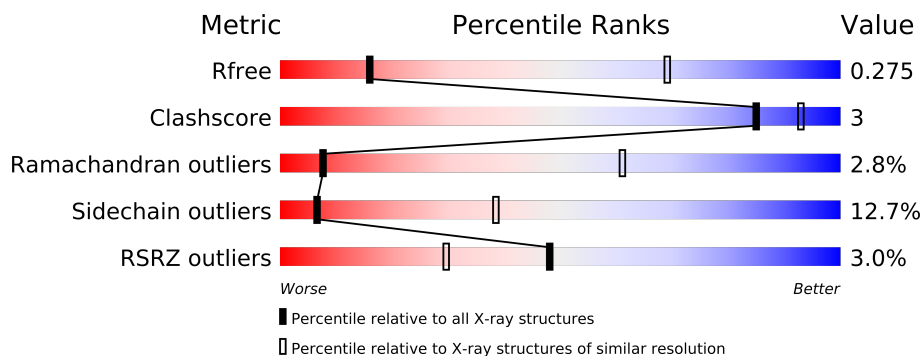
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable22639
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable22683

# 1 Overall quality at a glance

The reported resolution of this entry is 3.70 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))
$R_{free}$	66092	1098 (4.00-3.40)
Clashscore	79885	1009 (3.94-3.46)
Ramachandran outliers	78287	1016 (3.98-3.42)
Sidechain outliers	78261	1014 (3.98-3.42)
RSRZ outliers	66119	1099 (4.00-3.40)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	449	
1	B	449	
1	C	449	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
2	ZN	B	451	-	X

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 9211 atoms, of which 0 are hydrogen and 0 are deuterium.

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	405	Total	C	N	O	S	8	0	0
			3084	1992	502	582	8			
1	B	398	Total	C	N	O	S	4	0	0
			3049	1961	500	580	8			
1	C	402	Total	C	N	O	S	0	0	0
			3065	1971	503	583	8			

There are 111 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	359	ALA	-	LINKER	? ?
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	? ?

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	436	THR	-	EXPRESSION TAG	? ?
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	359	ALA	-	LINKER	? ?
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	? ?

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	441	GLY	-	EXPRESSION TAG	? ?
B	442	THR	-	EXPRESSION TAG	? ?
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	359	ALA	-	LINKER	? ?
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	? ?

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	446	PHE	-	EXPRESSION TAG	? ?
C	447	THR	-	EXPRESSION TAG	? ?
C	448	THR	-	EXPRESSION TAG	? ?
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 Zn 3 3	0	0
2	A	2	Total Zn 2 2	0	0
2	C	1	Total Zn 1 1	0	0

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

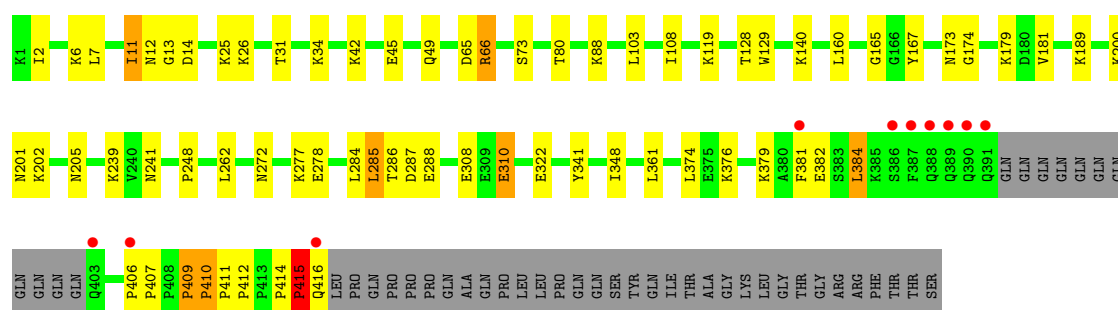
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	2	Total Ca 2 2	0	0
3	A	3	Total Ca 3 3	0	0
3	C	2	Total Ca 2 2	0	0

### 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.

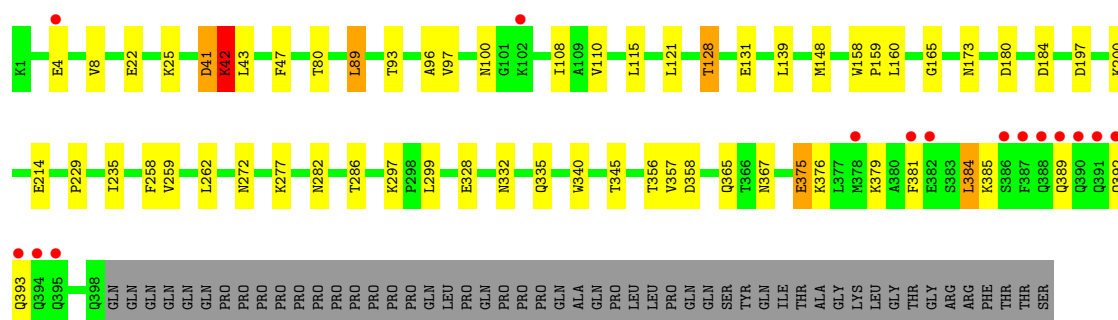
- Molecule 1: Maltose-binding protein, huntingtin fusion protein

Chain A:



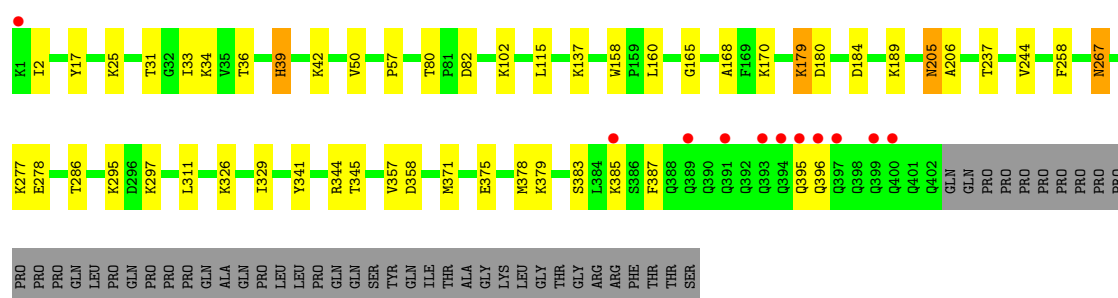
- Molecule 1: Maltose-binding protein, huntingtin fusion protein

Chain B:



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

Chain C:



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	163.82Å 101.19Å 134.82Å 90.00° 99.23° 90.00°	Depositor
Resolution (Å)	38.00 – 3.70 37.53 – 3.70	Depositor EDS
% Data completeness (in resolution range)	92.8 (38.00-3.70) 92.8 (37.53-3.70)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.08	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.18 (at 3.66Å)	Xtriage
Refinement program	REFMAC 5.5.0102	Depositor
R, $R_{free}$	0.264 , 0.280 0.261 , 0.275	Depositor DCC
$R_{free}$ test set	1095 reflections (5.31%)	DCC
Wilson B-factor (Å <sup>2</sup> )	85.4	Xtriage
Anisotropy	0.067	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , 22.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Outliers	0 of 21724 reflections	Xtriage
$F_o, F_c$ correlation	0.87	EDS
Total number of atoms	9211	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	84.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.52% of the height of the origin peak. No significant pseudotranslation is detected.*

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



## 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	0.42	2/3164 (0.1%)	0.52	1/4307 (0.0%)
1	B	0.43	1/3120 (0.0%)	0.51	1/4236 (0.0%)
1	C	0.37	0/3136	0.49	0/4259
All	All	0.40	3/9420 (0.0%)	0.51	2/12802 (0.0%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	328	GLU	CB-CG	-12.62	1.28	1.52
1	A	310	GLU	CB-CG	-9.75	1.33	1.52
1	A	6	LYS	CB-CG	-8.20	1.30	1.52

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	415	PRO	N-CA-CB	5.16	109.49	103.30
1	B	384	LEU	CA-CB-CG	5.14	127.13	115.30

There are no chirality outliers.

There are no planarity outliers.

### 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	3084	0	0	9	0
1	B	3049	0	0	11	0
1	C	3065	0	0	7	0
2	A	2	0	0	0	0
2	B	3	0	0	0	0
2	C	1	0	0	0	0
3	A	3	0	0	0	0
3	B	2	0	0	0	0
3	C	2	0	0	0	0
All	All	9211	0	0	26	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 3.

The worst 5 of 26 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:415:PRO:CB	1:A:416:GLN:CA	2.65	0.74
1:C:57:PRO:O	1:C:267:ASN:ND2	2.24	0.71
1:C:39:HIS:CG	1:C:39:HIS:O	2.44	0.69
1:A:410:PRO:CB	1:A:411:PRO:CD	2.74	0.66
1:A:286:THR:O	1:A:288:GLU:N	2.38	0.57

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone ⓘ

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	401/449 (89%)	346 (86%)	36 (9%)	19 (5%)	4 46
1	B	396/449 (88%)	364 (92%)	25 (6%)	7 (2%)	13 70
1	C	400/449 (89%)	350 (88%)	42 (10%)	8 (2%)	11 68
All	All	1197/1347 (89%)	1060 (89%)	103 (9%)	34 (3%)	8 61

5 of 34 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	2	ILE
1	A	14	ASP
1	A	181	VAL
1	A	285	LEU
1	A	287	ASP

### 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	A	311/363 (86%)	271 (87%)	40 (13%)	6	38
1	B	306/363 (84%)	267 (87%)	39 (13%)	6	38
1	C	305/363 (84%)	267 (88%)	38 (12%)	7	40
All	All	922/1089 (85%)	805 (87%)	117 (13%)	6	38

5 of 117 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	128	THR
1	B	282	ASN
1	C	329	ILE
1	B	148	MET
1	B	200	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

There are no RNA chains 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 carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 13 ligands modelled in this entry, 13 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.

## 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 ⓘ

### 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	A	405/449 (90%)	0.15	10 (2%) 54 34	58, 83, 93, 104	37 (9%)
1	B	398/449 (88%)	0.16	15 (3%) 38 25	72, 84, 94, 100	30 (7%)
1	C	402/449 (89%)	0.16	11 (2%) 52 32	74, 85, 98, 108	33 (8%)
All	All	1205/1347 (89%)	0.16	36 (2%) 48 29	58, 84, 96, 108	100 (8%)

The worst 5 of 36 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	387	PHE	4.4
1	B	391	GLN	4.0
1	B	390	GLN	3.9
1	B	388	GLN	3.9
1	B	382	GLU	3.7

### 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 carbohydrates in this entry.

### 6.4 Ligands ⓘ

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	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	ZN	B	451	1/1	0.41	2.02	99,99,99,99	0
3	CA	B	454	1/1	0.20	-0.02	80,80,80,80	0
3	CA	A	445	1/1	0.22	-0.38	58,58,58,58	0
3	CA	C	430	1/1	0.28	-0.61	94,94,94,94	0
3	CA	A	444	1/1	0.21	-0.75	54,54,54,54	0
2	ZN	A	442	1/1	0.18	-1.00	76,76,76,76	0
2	ZN	B	452	1/1	0.15	-1.22	65,65,65,65	0
3	CA	C	429	1/1	0.19	-1.23	96,96,96,96	0
3	CA	B	453	1/1	0.14	-2.02	74,74,74,74	0
3	CA	A	446	1/1	0.16	-2.81	89,89,89,89	0
2	ZN	B	450	1/1	0.12	-3.13	105,105,105,105	0
2	ZN	C	428	1/1	0.10	-5.56	110,110,110,110	0
2	ZN	A	443	1/1	0.12	-20.00	116,116,116,116	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.