wwPDB X-ray Structure Validation Summary Report (i)

Feb 13, 2024 – 05:24 AM EST

PDB ID	:	3IO6
Title	:	Huntingtin amino-terminal region with 17 Gln residues - crystal C92-a
Authors	:	Kim, M.W.; Chelliah, Y.; Kim, S.W.; Otwinowski, Z.; Bezprozvanny, I.
Deposited on	:	2009-08-13
Resolution	:	3.70 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) 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 (i)) were used in the production of this report:

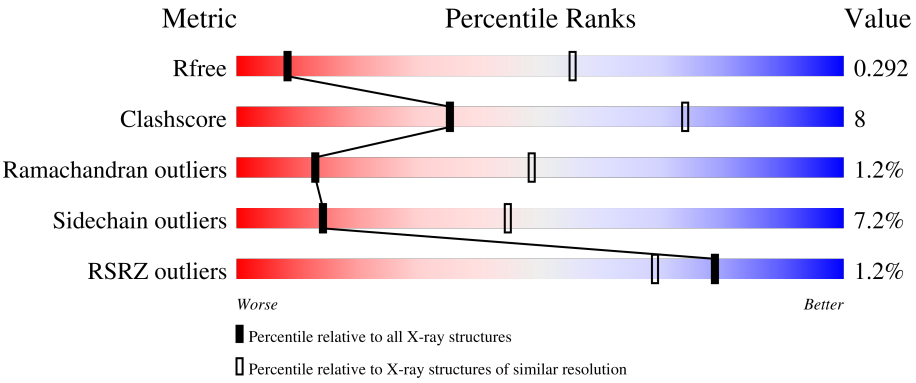
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 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}	130704	1049 (3.88-3.52)
Clashscore	141614	1027 (3.86-3.54)
Ramachandran outliers	138981	1069 (3.88-3.52)
Sidechain outliers	138945	1065 (3.88-3.52)
RSRZ outliers	127900	1578 (3.90-3.50)

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 >=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 <=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	71% 17% • 11%
1	B	449	% 75% 13% • 11%
1	C	449	2% 68% 18% • 12%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 9265 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 periplasmic protein, HUNTINGTIN FUSION PROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	400	Total	C	N	O	S	8	0	0
			3059	1969	499	583	8			
1	C	397	Total	C	N	O	S	4	0	0
			3076	1976	504	588	8			
1	B	401	Total	C	N	O	S	20	0	0
			3116	1998	513	597	8			

There are 105 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	359	ALA	-	linker	UNP P0AEY0
A	360	ALA	-	linker	UNP P0AEY0
A	361	LEU	-	linker	UNP P0AEY0
A	362	ALA	-	linker	UNP P0AEY0
A	363	ALA	-	linker	UNP P0AEY0
A	364	ALA	-	linker	UNP P0AEY0
A	365	GLN	-	linker	UNP P0AEY0
A	366	THR	-	linker	UNP P0AEY0
A	367	ASN	-	linker	UNP P0AEY0
A	368	ALA	-	linker	UNP P0AEY0
A	369	ALA	-	linker	UNP P0AEY0
A	370	ALA	-	linker	UNP P0AEY0
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	UNP P42858
A	432	SER	-	expression tag	UNP P42858
A	433	TYR	-	expression tag	UNP P42858
A	434	GLN	-	expression tag	UNP P42858
A	435	ILE	-	expression tag	UNP P42858
A	436	THR	-	expression tag	UNP P42858

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	437	ALA	-	expression tag	UNP P42858
A	438	GLY	-	expression tag	UNP P42858
A	439	LYS	-	expression tag	UNP P42858
A	440	LEU	-	expression tag	UNP P42858
A	441	GLY	-	expression tag	UNP P42858
A	442	THR	-	expression tag	UNP P42858
A	443	GLY	-	expression tag	UNP P42858
A	444	ARG	-	expression tag	UNP P42858
A	445	ARG	-	expression tag	UNP P42858
A	446	PHE	-	expression tag	UNP P42858
A	447	THR	-	expression tag	UNP P42858
A	448	THR	-	expression tag	UNP P42858
A	449	SER	-	expression tag	UNP P42858
C	359	ALA	-	linker	UNP P0AEY0
C	360	ALA	-	linker	UNP P0AEY0
C	361	LEU	-	linker	UNP P0AEY0
C	362	ALA	-	linker	UNP P0AEY0
C	363	ALA	-	linker	UNP P0AEY0
C	364	ALA	-	linker	UNP P0AEY0
C	365	GLN	-	linker	UNP P0AEY0
C	366	THR	-	linker	UNP P0AEY0
C	367	ASN	-	linker	UNP P0AEY0
C	368	ALA	-	linker	UNP P0AEY0
C	369	ALA	-	linker	UNP P0AEY0
C	370	ALA	-	linker	UNP P0AEY0
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	UNP P42858
C	432	SER	-	expression tag	UNP P42858
C	433	TYR	-	expression tag	UNP P42858
C	434	GLN	-	expression tag	UNP P42858
C	435	ILE	-	expression tag	UNP P42858
C	436	THR	-	expression tag	UNP P42858
C	437	ALA	-	expression tag	UNP P42858
C	438	GLY	-	expression tag	UNP P42858
C	439	LYS	-	expression tag	UNP P42858
C	440	LEU	-	expression tag	UNP P42858
C	441	GLY	-	expression tag	UNP P42858
C	442	THR	-	expression tag	UNP P42858
C	443	GLY	-	expression tag	UNP P42858

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	444	ARG	-	expression tag	UNP P42858
C	445	ARG	-	expression tag	UNP P42858
C	446	PHE	-	expression tag	UNP P42858
C	447	THR	-	expression tag	UNP P42858
C	448	THR	-	expression tag	UNP P42858
C	449	SER	-	expression tag	UNP P42858
B	359	ALA	-	linker	UNP P0AEY0
B	360	ALA	-	linker	UNP P0AEY0
B	361	LEU	-	linker	UNP P0AEY0
B	362	ALA	-	linker	UNP P0AEY0
B	363	ALA	-	linker	UNP P0AEY0
B	364	ALA	-	linker	UNP P0AEY0
B	365	GLN	-	linker	UNP P0AEY0
B	366	THR	-	linker	UNP P0AEY0
B	367	ASN	-	linker	UNP P0AEY0
B	368	ALA	-	linker	UNP P0AEY0
B	369	ALA	-	linker	UNP P0AEY0
B	370	ALA	-	linker	UNP P0AEY0
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	UNP P42858
B	432	SER	-	expression tag	UNP P42858
B	433	TYR	-	expression tag	UNP P42858
B	434	GLN	-	expression tag	UNP P42858
B	435	ILE	-	expression tag	UNP P42858
B	436	THR	-	expression tag	UNP P42858
B	437	ALA	-	expression tag	UNP P42858
B	438	GLY	-	expression tag	UNP P42858
B	439	LYS	-	expression tag	UNP P42858
B	440	LEU	-	expression tag	UNP P42858
B	441	GLY	-	expression tag	UNP P42858
B	442	THR	-	expression tag	UNP P42858
B	443	GLY	-	expression tag	UNP P42858
B	444	ARG	-	expression tag	UNP P42858
B	445	ARG	-	expression tag	UNP P42858
B	446	PHE	-	expression tag	UNP P42858
B	447	THR	-	expression tag	UNP P42858
B	448	THR	-	expression tag	UNP P42858
B	449	SER	-	expression tag	UNP P42858

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

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total Zn 1 1	0	0
2	C	2	Total Zn 2 2	0	0
2	B	2	Total Zn 2 2	0	0

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

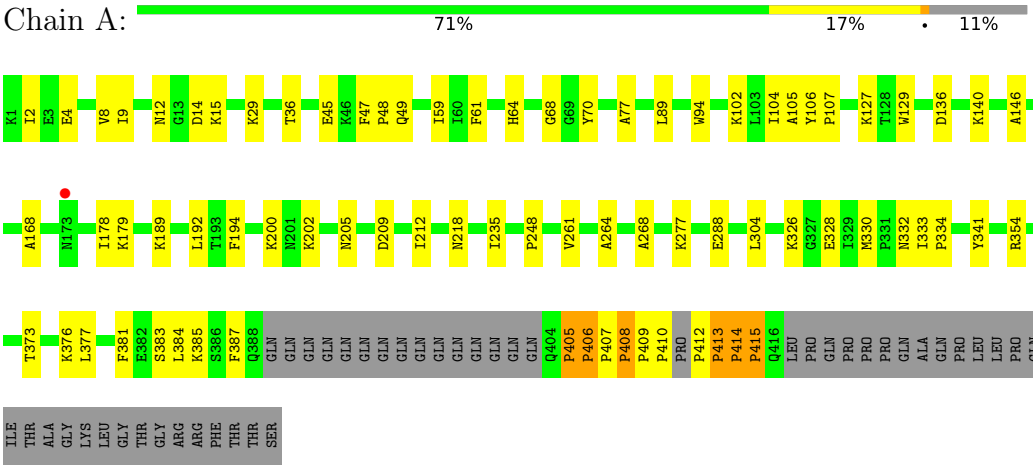
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	3	Total Ca 3 3	0	0
3	C	4	Total Ca 4 4	0	0
3	B	2	Total Ca 2 2	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

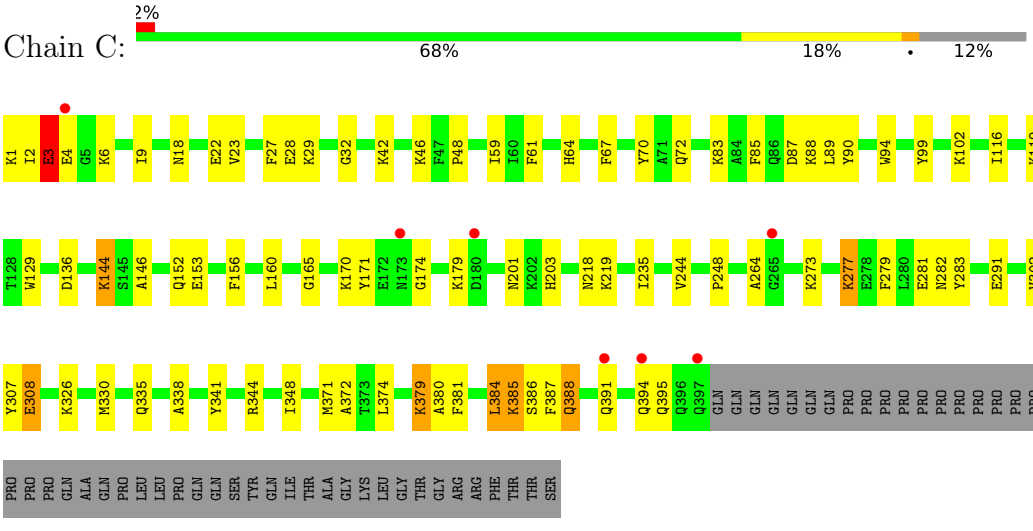
• Molecule 1: Maltose-binding periplasmic protein, HUNTINGTIN FUSION PROTEIN

Chain A:



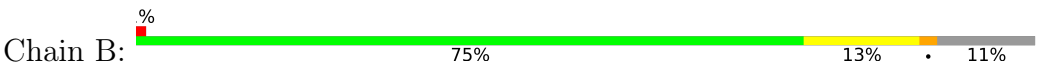
• Molecule 1: Maltose-binding periplasmic protein, HUNTINGTIN FUSION PROTEIN

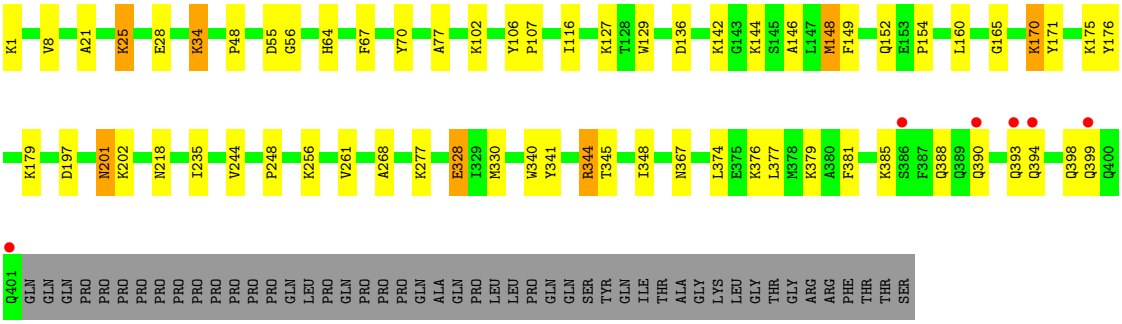
Chain C:



• Molecule 1: Maltose-binding periplasmic protein, HUNTINGTIN FUSION PROTEIN

Chain B:




K1 V8 A21 K25 E28 K34 P48 D55 G56 H64 F67 Y70 A77 K102 Y106 P107 I116 K127 T128 W129 D136 K142 G143 K144 S145 A146 L147 M148 F149 Q152 E153 P154 L160 G165 K170 Y171 K175 Y176
K179 D197 N201 K202 N218 I235 V244 P248 K256 V261 A268 K277 E328 I329 M330 W340 Y341 R344 T345 I348 N367 L374 E375 K376 L377 M378 K379 A380 F381 K385 S386 F387 Q388 Q389 Q390 Q393 Q394 Q398 Q399 Q400
Q401 GLN GLN GLN PRO PRO PRO PRO PRO PRO PRO PRO PRO PRO PRO PRO GLN LEU PRO GLN PRO PRO PRO GLN ALA GLN PRO LEU LEU PRO GLN GLN SER TYR GLN ILE THR ALA GLY LYS LEU GLY THR GLY ARG ARG PHE THR THR SER

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	163.40Å 101.24Å 138.19Å 90.00° 92.00° 90.00°	Depositor
Resolution (Å)	38.67 – 3.70 29.73 – 3.70	Depositor EDS
% Data completeness (in resolution range)	92.4 (38.67-3.70) 92.5 (29.73-3.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.02	Depositor
$< I/\sigma(I) >$ [1]	2.23 (at 3.75Å)	Xtriage
Refinement program	REFMAC 5.5.0102	Depositor
R, R_{free}	0.245 , 0.293 0.247 , 0.292	Depositor DCC
R_{free} test set	1200 reflections (5.36%)	wwPDB-VP
Wilson B-factor (Å^2)	98.6	Xtriage
Anisotropy	0.234	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.26 , 39.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.012 for -h,-k,l	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	9265	wwPDB-VP
Average B, all atoms (Å^2)	81.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.01% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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: CA, ZN

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	0.40	1/3128 (0.0%)	0.57	10/4243 (0.2%)
1	B	0.46	2/3187 (0.1%)	0.60	10/4321 (0.2%)
1	C	0.43	1/3147 (0.0%)	0.51	2/4268 (0.0%)
All	All	0.43	4/9462 (0.0%)	0.56	22/12832 (0.2%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	256	LYS	CB-CG	-13.33	1.16	1.52
1	C	144	LYS	CB-CG	-12.71	1.18	1.52
1	A	140	LYS	CB-CG	-10.35	1.24	1.52
1	B	277	LYS	CB-CG	9.03	1.76	1.52

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	144	LYS	CA-CB-CG	9.46	134.20	113.40
1	B	34	LYS	CA-CB-CG	-9.24	93.07	113.40
1	B	34	LYS	CB-CG-CD	-8.77	88.79	111.60
1	B	256	LYS	CA-CB-CG	-7.62	96.64	113.40
1	B	277	LYS	CA-CB-CG	-7.27	97.41	113.40

There are no chirality outliers.

There are no planarity outliers.

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	3059	0	2996	49	0
1	B	3116	0	3087	39	0
1	C	3076	0	3049	84	0
2	A	1	0	0	0	0
2	B	2	0	0	0	0
2	C	2	0	0	0	0
3	A	3	0	0	0	0
3	B	2	0	0	0	0
3	C	4	0	0	0	0
All	All	9265	0	9132	150	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 150 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:A:387:PHE:HB3	1:B:341:TYR:OH	1.30	1.27
1:A:341:TYR:CE2	1:C:387:PHE:CE1	2.27	1.22
1:C:385:LYS:HD2	1:C:385:LYS:O	1.42	1.18
1:A:341:TYR:CZ	1:C:387:PHE:CZ	2.34	1.14
1:A:341:TYR:CE2	1:C:387:PHE:CZ	2.42	1.07

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	394/449 (88%)	367 (93%)	19 (5%)	8 (2%)	7 39

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	B	399/449 (89%)	377 (94%)	21 (5%)	1 (0%)	41 74
1	C	395/449 (88%)	361 (91%)	29 (7%)	5 (1%)	12 47
All	All	1188/1347 (88%)	1105 (93%)	69 (6%)	14 (1%)	13 48

5 of 14 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	405	PRO
1	A	406	PRO
1	A	408	PRO
1	A	413	PRO
1	A	414	PRO

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	307/363 (85%)	288 (94%)	19 (6%)	18 49
1	B	320/363 (88%)	297 (93%)	23 (7%)	14 45
1	C	315/363 (87%)	289 (92%)	26 (8%)	11 40
All	All	942/1089 (86%)	874 (93%)	68 (7%)	14 45

5 of 68 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	179	LYS
1	B	202	LYS
1	B	390	GLN
1	C	83	LYS
1	C	42	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 23 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	394	GLN
1	B	64	HIS
1	B	49	GLN
1	B	124	ASN
1	A	282	ASN

5.3.3 RNA

There are no RNA molecules 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 monosaccharides in this entry.

5.6 Ligand geometry

Of 14 ligands modelled in this entry, 14 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

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^{th} 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^2)	Q<0.9
1	A	400/449 (89%)	-0.36	1 (0%)	94	90	80, 80, 108, 108	27 (6%)
1	B	401/449 (89%)	-0.27	6 (1%)	73	63	80, 80, 108, 108	33 (8%)
1	C	397/449 (88%)	-0.32	7 (1%)	68	57	80, 80, 80, 108	23 (5%)
All	All	1198/1347 (88%)	-0.31	14 (1%)	79	69	80, 80, 108, 108	83 (6%)

The worst 5 of 14 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	386	SER	3.7
1	B	390	GLN	3.6
1	C	391	GLN	3.2
1	C	173	ASN	3.2
1	B	399	GLN	3.0

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 monosaccharides 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. The B-factors column lists the minimum, median, 95^{th} 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(Å^2)	Q<0.9
3	CA	B	452	1/1	0.53	0.20	107,107,107,107	0
2	ZN	C	451	1/1	0.81	0.15	107,107,107,107	0
3	CA	B	453	1/1	0.82	0.28	107,107,107,107	0
3	CA	C	454	1/1	0.85	0.17	107,107,107,107	0
3	CA	A	453	1/1	0.86	0.26	107,107,107,107	0
3	CA	C	455	1/1	0.87	0.13	107,107,107,107	0
3	CA	C	452	1/1	0.88	0.19	107,107,107,107	0
2	ZN	B	450	1/1	0.91	0.09	107,107,107,107	0
2	ZN	A	450	1/1	0.95	0.10	107,107,107,107	0
2	ZN	C	450	1/1	0.96	0.22	107,107,107,107	0
3	CA	A	451	1/1	0.96	0.22	107,107,107,107	0
3	CA	A	452	1/1	0.97	0.10	107,107,107,107	0
3	CA	C	453	1/1	0.97	0.26	107,107,107,107	0
2	ZN	B	451	1/1	0.97	0.08	107,107,107,107	0

6.5 Other polymers (i)

There are no such residues in this entry.