wwPDB X-ray Structure Validation Summary Report (i)

Mar 9, 2018 – 01:56 pm GMT

PDB ID	:	2HYX
Title	:	Structure of the C-terminal domain of DipZ from Mycobacterium tuberculosis
Authors	:	Goldstone, D.; Baker, E.N.; Metcalf, P.; TB Structural Genomics Consortium (TBSGC)
Deposited on	:	2006-08-08
Resolution	:	1.90 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 following versions of software and data (see references (1)) were used in the production of this report:

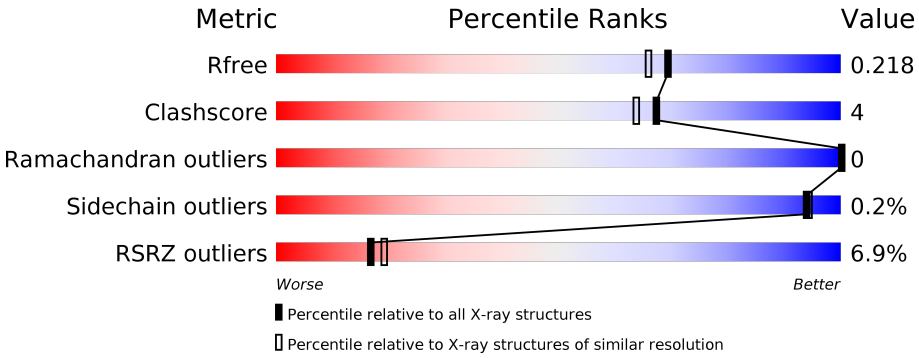
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	trunk30967
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac	:	5.8.0158
CCP4	:	7.0 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk30967

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 1.90 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}	111664	5502 (1.90-1.90)
Clashscore	122126	6115 (1.90-1.90)
Ramachandran outliers	120053	6048 (1.90-1.90)
Sidechain outliers	120020	6048 (1.90-1.90)
RSRZ outliers	108989	5379 (1.90-1.90)

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 >=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 <=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	352	5% 84% 10% 5%
1	B	352	14% 81% 13% 7%
1	C	352	4% 87% 7% 5%
1	D	352	3% 86% 7% 7%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 10818 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 dipZ.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	333	Total	C	N	O	S	0	0	0
			2503	1592	425	481	5			
1	B	328	Total	C	N	O	S	0	0	0
			2434	1549	410	470	5			
1	C	333	Total	C	N	O	S	0	0	0
			2501	1590	425	481	5			
1	D	329	Total	C	N	O	S	0	0	0
			2481	1578	424	474	5			

There are 88 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-22	MET	-	CLONING ARTIFACT	UNP Q10801
A	-21	ARG	-	CLONING ARTIFACT	UNP Q10801
A	-20	GLY	-	CLONING ARTIFACT	UNP Q10801
A	-19	SER	-	CLONING ARTIFACT	UNP Q10801
A	-18	HIS	-	CLONING ARTIFACT	UNP Q10801
A	-17	HIS	-	CLONING ARTIFACT	UNP Q10801
A	-16	HIS	-	CLONING ARTIFACT	UNP Q10801
A	-15	HIS	-	CLONING ARTIFACT	UNP Q10801
A	-14	HIS	-	CLONING ARTIFACT	UNP Q10801
A	-13	HIS	-	CLONING ARTIFACT	UNP Q10801
A	-12	GLY	-	CLONING ARTIFACT	UNP Q10801
A	-11	SER	-	CLONING ARTIFACT	UNP Q10801
A	-10	GLU	-	CLONING ARTIFACT	UNP Q10801
A	-9	ASN	-	CLONING ARTIFACT	UNP Q10801
A	-8	LEU	-	CLONING ARTIFACT	UNP Q10801
A	-7	TYR	-	CLONING ARTIFACT	UNP Q10801
A	-6	PHE	-	CLONING ARTIFACT	UNP Q10801
A	-5	GLN	-	CLONING ARTIFACT	UNP Q10801
A	-4	SER	-	CLONING ARTIFACT	UNP Q10801
A	-3	GLY	-	CLONING ARTIFACT	UNP Q10801
A	-2	ALA	-	CLONING ARTIFACT	UNP Q10801

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MET	-	CLONING ARTIFACT	UNP Q10801
B	-22	MET	-	CLONING ARTIFACT	UNP Q10801
B	-21	ARG	-	CLONING ARTIFACT	UNP Q10801
B	-20	GLY	-	CLONING ARTIFACT	UNP Q10801
B	-19	SER	-	CLONING ARTIFACT	UNP Q10801
B	-18	HIS	-	CLONING ARTIFACT	UNP Q10801
B	-17	HIS	-	CLONING ARTIFACT	UNP Q10801
B	-16	HIS	-	CLONING ARTIFACT	UNP Q10801
B	-15	HIS	-	CLONING ARTIFACT	UNP Q10801
B	-14	HIS	-	CLONING ARTIFACT	UNP Q10801
B	-13	HIS	-	CLONING ARTIFACT	UNP Q10801
B	-12	GLY	-	CLONING ARTIFACT	UNP Q10801
B	-11	SER	-	CLONING ARTIFACT	UNP Q10801
B	-10	GLU	-	CLONING ARTIFACT	UNP Q10801
B	-9	ASN	-	CLONING ARTIFACT	UNP Q10801
B	-8	LEU	-	CLONING ARTIFACT	UNP Q10801
B	-7	TYR	-	CLONING ARTIFACT	UNP Q10801
B	-6	PHE	-	CLONING ARTIFACT	UNP Q10801
B	-5	GLN	-	CLONING ARTIFACT	UNP Q10801
B	-4	SER	-	CLONING ARTIFACT	UNP Q10801
B	-3	GLY	-	CLONING ARTIFACT	UNP Q10801
B	-2	ALA	-	CLONING ARTIFACT	UNP Q10801
B	-1	MET	-	CLONING ARTIFACT	UNP Q10801
C	-22	MET	-	CLONING ARTIFACT	UNP Q10801
C	-21	ARG	-	CLONING ARTIFACT	UNP Q10801
C	-20	GLY	-	CLONING ARTIFACT	UNP Q10801
C	-19	SER	-	CLONING ARTIFACT	UNP Q10801
C	-18	HIS	-	CLONING ARTIFACT	UNP Q10801
C	-17	HIS	-	CLONING ARTIFACT	UNP Q10801
C	-16	HIS	-	CLONING ARTIFACT	UNP Q10801
C	-15	HIS	-	CLONING ARTIFACT	UNP Q10801
C	-14	HIS	-	CLONING ARTIFACT	UNP Q10801
C	-13	HIS	-	CLONING ARTIFACT	UNP Q10801
C	-12	GLY	-	CLONING ARTIFACT	UNP Q10801
C	-11	SER	-	CLONING ARTIFACT	UNP Q10801
C	-10	GLU	-	CLONING ARTIFACT	UNP Q10801
C	-9	ASN	-	CLONING ARTIFACT	UNP Q10801
C	-8	LEU	-	CLONING ARTIFACT	UNP Q10801
C	-7	TYR	-	CLONING ARTIFACT	UNP Q10801
C	-6	PHE	-	CLONING ARTIFACT	UNP Q10801
C	-5	GLN	-	CLONING ARTIFACT	UNP Q10801
C	-4	SER	-	CLONING ARTIFACT	UNP Q10801

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	-3	GLY	-	CLONING ARTIFACT	UNP Q10801
C	-2	ALA	-	CLONING ARTIFACT	UNP Q10801
C	-1	MET	-	CLONING ARTIFACT	UNP Q10801
D	-22	MET	-	CLONING ARTIFACT	UNP Q10801
D	-21	ARG	-	CLONING ARTIFACT	UNP Q10801
D	-20	GLY	-	CLONING ARTIFACT	UNP Q10801
D	-19	SER	-	CLONING ARTIFACT	UNP Q10801
D	-18	HIS	-	CLONING ARTIFACT	UNP Q10801
D	-17	HIS	-	CLONING ARTIFACT	UNP Q10801
D	-16	HIS	-	CLONING ARTIFACT	UNP Q10801
D	-15	HIS	-	CLONING ARTIFACT	UNP Q10801
D	-14	HIS	-	CLONING ARTIFACT	UNP Q10801
D	-13	HIS	-	CLONING ARTIFACT	UNP Q10801
D	-12	GLY	-	CLONING ARTIFACT	UNP Q10801
D	-11	SER	-	CLONING ARTIFACT	UNP Q10801
D	-10	GLU	-	CLONING ARTIFACT	UNP Q10801
D	-9	ASN	-	CLONING ARTIFACT	UNP Q10801
D	-8	LEU	-	CLONING ARTIFACT	UNP Q10801
D	-7	TYR	-	CLONING ARTIFACT	UNP Q10801
D	-6	PHE	-	CLONING ARTIFACT	UNP Q10801
D	-5	GLN	-	CLONING ARTIFACT	UNP Q10801
D	-4	SER	-	CLONING ARTIFACT	UNP Q10801
D	-3	GLY	-	CLONING ARTIFACT	UNP Q10801
D	-2	ALA	-	CLONING ARTIFACT	UNP Q10801
D	-1	MET	-	CLONING ARTIFACT	UNP Q10801

- Molecule 2 is water.

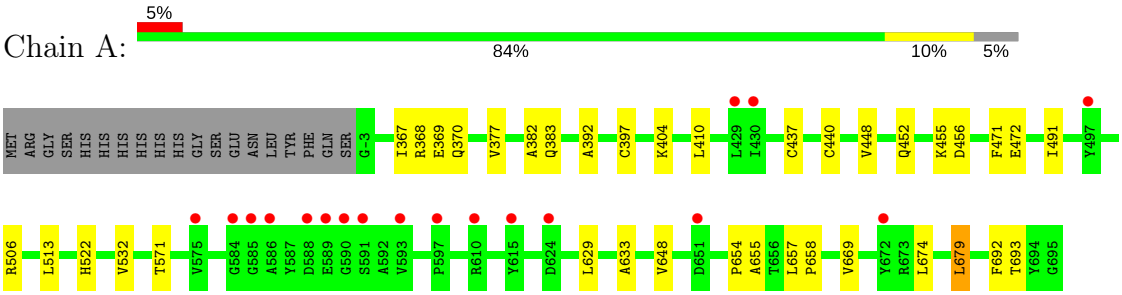
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	228	Total O 228 228	0	0
2	B	167	Total O 167 167	0	0
2	C	247	Total O 247 247	0	0
2	D	257	Total O 257 257	0	0

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 geometryand 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.

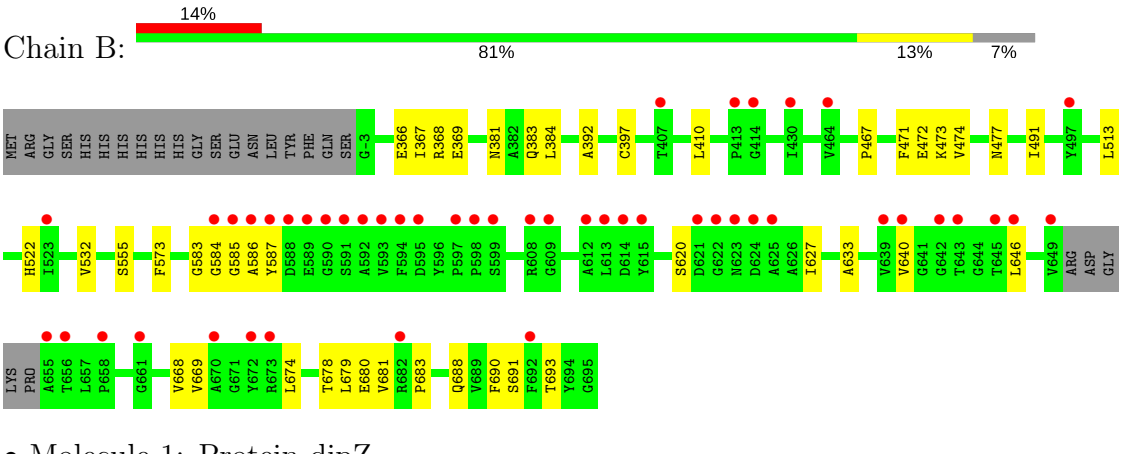
• Molecule 1: Protein dipZ

Chain A:



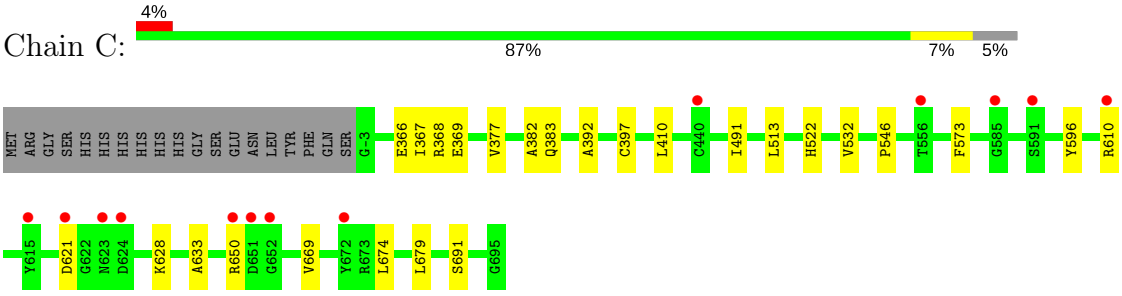
• Molecule 1: Protein dipZ

Chain B:



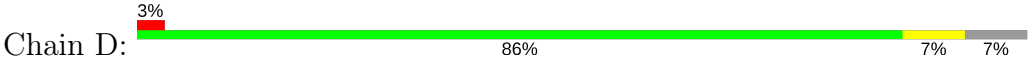
• Molecule 1: Protein dipZ

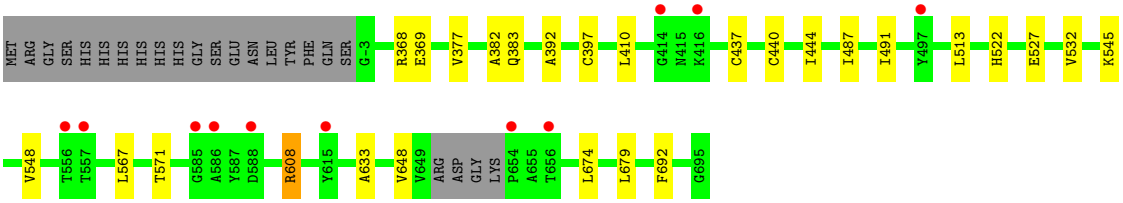
Chain C:



• Molecule 1: Protein dipZ

Chain D:




MET
ARG
GLY
SER
HIS
HIS
HIS
HIS
HIS
HIS
GLY
SER
GLU
ASN
LEU
TYR
PHE
GLN
SER
G-3
R368
E369
V377
A382
Q383
A392
C397
L410
G414
N415
K416
C437
C440
I444
I487
I491
Y497
L513
H522
E527
V532
K545
V548
T556
T557
L567
T571
G585
A586
Y587
D588
R608
Y615
A633
V648
V649
ARG
ASP
GLY
LYS
P654
A655
T656
L674
L679
F692
G695

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	110.07Å 117.92Å 123.03Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 1.90 29.76 – 1.90	Depositor EDS
% Data completeness (in resolution range)	97.6 (30.00-1.90) 97.1 (29.76-1.90)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.84 (at 1.89Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.195 , 0.219 0.195 , 0.218	Depositor DCC
R_{free} test set	12569 reflections (9.96%)	wwPDB-VP
Wilson B-factor (Å^2)	22.5	Xtriage
Anisotropy	0.280	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 50.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.012 for -h,l,k	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	10818	wwPDB-VP
Average B, all atoms (Å^2)	24.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.73% 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

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	0.32	0/2566	0.63	0/3507
1	B	0.61	9/2494 (0.4%)	0.64	0/3415
1	C	0.31	0/2564	0.63	0/3505
1	D	0.32	0/2543	0.64	0/3473
All	All	0.41	9/10167 (0.1%)	0.63	0/13900

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	587	TYR	CE1-CZ	-10.81	1.24	1.38
1	B	586	ALA	CA-CB	-8.79	1.33	1.52
1	B	587	TYR	CG-CD1	-8.43	1.28	1.39
1	B	587	TYR	CE2-CZ	-7.52	1.28	1.38
1	B	587	TYR	CG-CD2	-7.07	1.29	1.39

There are no bond angle outliers.

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	2503	0	2434	25	0
1	B	2434	0	2331	25	0
1	C	2501	0	2430	16	0
1	D	2481	0	2421	17	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	228	0	0	2	0
2	B	167	0	0	2	0
2	C	247	0	0	1	0
2	D	257	0	0	1	0
All	All	10818	0	9616	82	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 4.

The worst 5 of 82 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:610:ARG:HG2	1:C:621:ASP:HB3	1.72	0.70
1:D:633:ALA:HA	1:D:674:LEU:CD1	2.23	0.69
1:A:448:VAL:O	1:A:452:GLN:HG3	1.95	0.66
1:C:383:GLN:O	1:C:397:CYS:HB2	2.00	0.62
1:B:679:LEU:HD12	1:B:679:LEU:C	2.20	0.62

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	331/352 (94%)	322 (97%)	9 (3%)	0	100	100
1	B	324/352 (92%)	310 (96%)	14 (4%)	0	100	100
1	C	331/352 (94%)	321 (97%)	10 (3%)	0	100	100
1	D	325/352 (92%)	316 (97%)	9 (3%)	0	100	100
All	All	1311/1408 (93%)	1269 (97%)	42 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	257/279 (92%)	256 (100%)	1 (0%)	92 92
1	B	245/279 (88%)	245 (100%)	0	100 100
1	C	257/279 (92%)	257 (100%)	0	100 100
1	D	256/279 (92%)	255 (100%)	1 (0%)	92 92
All	All	1015/1116 (91%)	1013 (100%)	2 (0%)	94 94

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	679	LEU
1	D	608	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	393	GLN
1	C	452	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

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

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	333/352 (94%)	0.30	18 (5%)	26	29	11, 23, 35, 41	0
1	B	328/352 (93%)	0.81	49 (14%)	2	2	15, 28, 40, 44	0
1	C	333/352 (94%)	0.17	13 (3%)	39	43	12, 21, 34, 43	0
1	D	329/352 (93%)	0.21	11 (3%)	46	50	12, 22, 34, 40	0
All	All	1323/1408 (93%)	0.37	91 (6%)	17	19	11, 23, 37, 44	0

The worst 5 of 91 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	593	VAL	6.2
1	B	584	GLY	6.0
1	A	615	TYR	5.2
1	B	591	SER	5.1
1	B	612	ALA	4.9

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

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.