Full wwPDB X-ray Structure Validation Report ⓘ

Aug 22, 2023 – 10:08 AM EDT

PDB ID	:	8E67
Title	:	ETV6 H396Y variant bound to DNA containing the sequence GGAT
Authors	:	Scheu, K.; Chan, A.C.; Murphy, M.E.; McIntosh, L.P.
Deposited on	:	2022-08-22
Resolution	:	2.30 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 ⓘ 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 ⓘ) were used in the production of this report:

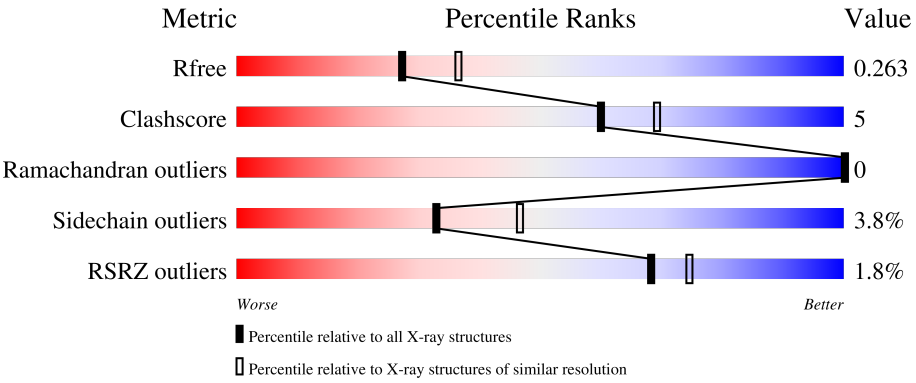
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 2.30 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	5042 (2.30-2.30)
Clashscore	141614	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)
RSRZ outliers	127900	4938 (2.30-2.30)

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	B	15	7% 67% 33%
1	D	15	73% 27%
1	G	15	47% 40% 13%
1	J	15	7% 73% 27%
2	C	15	73% 27%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	E	15	73% 27%
2	H	15	53% 40% 7%
2	K	15	80% 20%
3	A	101	% 72% 20% 8%
3	F	101	89% 6%
3	I	101	3% 75% 18% 7%
3	L	101	3% 77% 14% 9%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 5655 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 DNA chain called GGAT-containing 15 bp DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	15	Total	C	N	O	P	0	0	0
			315	148	65	87	15			
1	D	15	Total	C	N	O	P	0	0	0
			315	148	65	87	15			
1	G	13	Total	C	N	O	P	0	0	0
			273	128	55	77	13			
1	J	15	Total	C	N	O	P	0	0	0
			315	148	65	87	15			

- Molecule 2 is a DNA chain called Complementary 15 bp strand.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	15	Total	C	N	O	P	0	0	0
			300	144	48	93	15			
2	E	15	Total	C	N	O	P	0	0	0
			296	144	48	90	14			
2	H	14	Total	C	N	O	P	0	0	0
			280	134	46	86	14			
2	K	15	Total	C	N	O	P	0	0	0
			296	144	48	90	14			

- Molecule 3 is a protein called Transcription factor ETV6.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	93	Total	C	N	O	S	0	0	0
			809	521	146	138	4			
3	F	97	Total	C	N	O	S	0	0	0
			827	531	147	145	4			
3	I	94	Total	C	N	O	S	0	0	0
			813	523	144	142	4			
3	L	92	Total	C	N	O	S	0	0	0
			784	506	136	138	4			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	326	GLY	-	expression tag	UNP E9Q8J8
A	327	SER	-	expression tag	UNP E9Q8J8
A	328	MET	-	expression tag	UNP E9Q8J8
A	334	SER	CYS	conflict	UNP E9Q8J8
A	396	TYR	HIS	engineered mutation	UNP E9Q8J8
F	326	GLY	-	expression tag	UNP E9Q8J8
F	327	SER	-	expression tag	UNP E9Q8J8
F	328	MET	-	expression tag	UNP E9Q8J8
F	334	SER	CYS	conflict	UNP E9Q8J8
F	396	TYR	HIS	engineered mutation	UNP E9Q8J8
I	326	GLY	-	expression tag	UNP E9Q8J8
I	327	SER	-	expression tag	UNP E9Q8J8
I	328	MET	-	expression tag	UNP E9Q8J8
I	334	SER	CYS	conflict	UNP E9Q8J8
I	396	TYR	HIS	engineered mutation	UNP E9Q8J8
L	326	GLY	-	expression tag	UNP E9Q8J8
L	327	SER	-	expression tag	UNP E9Q8J8
L	328	MET	-	expression tag	UNP E9Q8J8
L	334	SER	CYS	conflict	UNP E9Q8J8
L	396	TYR	HIS	engineered mutation	UNP E9Q8J8

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	C	1	Total O 1 1	0	0
4	A	12	Total O 12 12	0	0
4	D	2	Total O 2 2	0	0
4	E	2	Total O 2 2	0	0
4	F	9	Total O 9 9	0	0
4	I	2	Total O 2 2	0	0
4	J	1	Total O 1 1	0	0
4	K	1	Total O 1 1	0	0
4	L	2	Total O 2 2	0	0

3 Residue-property plots (i)

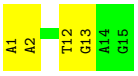
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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: GGAT-containing 15 bp DNA

Chain B:

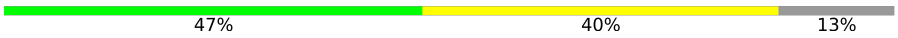
• Molecule 1: GGAT-containing 15 bp DNA

Chain D:



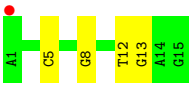
• Molecule 1: GGAT-containing 15 bp DNA

Chain G:



• Molecule 1: GGAT-containing 15 bp DNA

Chain J:

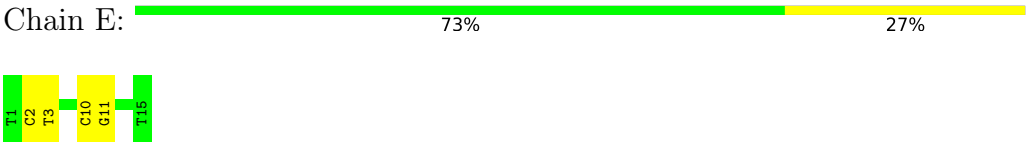


• Molecule 2: Complementary 15 bp strand

Chain C:

• Molecule 2: Complementary 15 bp strand

Chain E:

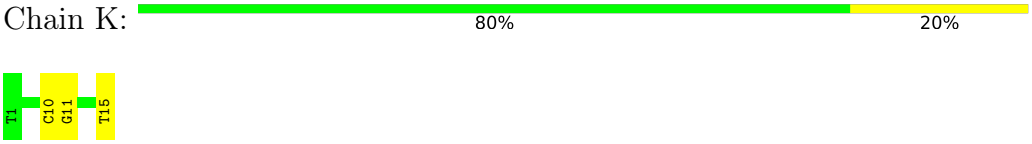


• Molecule 2: Complementary 15 bp strand

Chain H:

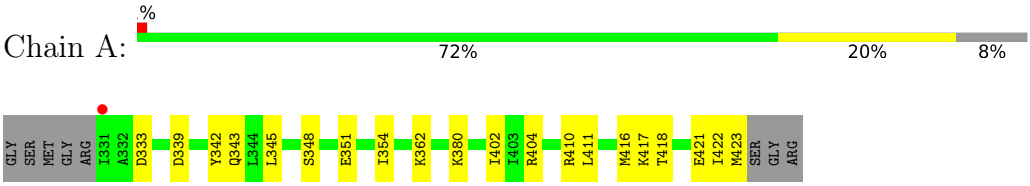
• Molecule 2: Complementary 15 bp strand

Chain K:



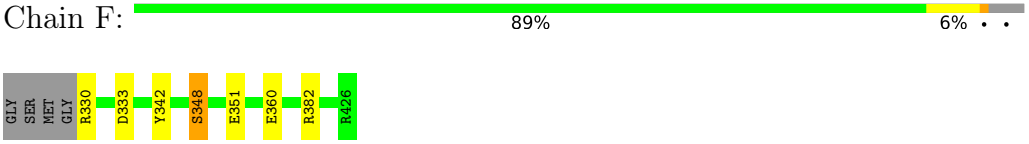
• Molecule 3: Transcription factor ETV6

Chain A:



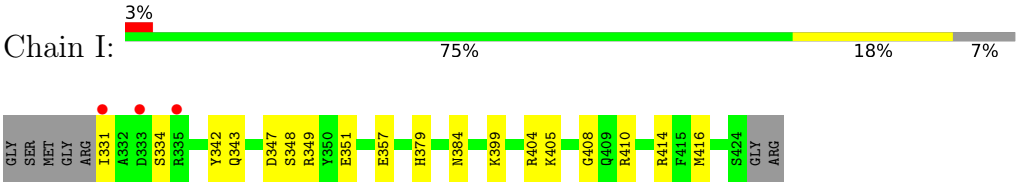
• Molecule 3: Transcription factor ETV6

Chain F:



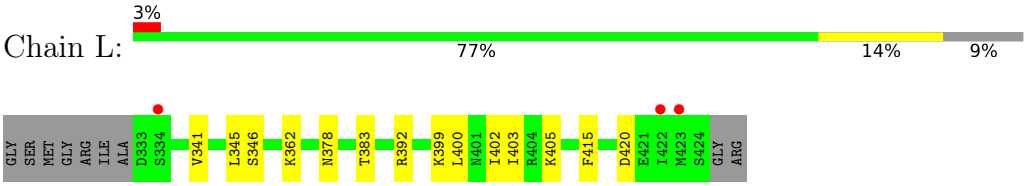
• Molecule 3: Transcription factor ETV6

Chain I:



• Molecule 3: Transcription factor ETV6

Chain L:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	63.94Å 74.35Å 76.62Å 90.00° 98.02° 90.00°	Depositor
Resolution (Å)	48.20 – 2.30 48.20 – 2.30	Depositor EDS
% Data completeness (in resolution range)	93.5 (48.20-2.30) 91.5 (48.20-2.30)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.23 (at 2.29Å)	Xtriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, R_{free}	0.211 , 0.267 0.213 , 0.263	Depositor DCC
R_{free} test set	1579 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	46.0	Xtriage
Anisotropy	0.166	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 44.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	5655	wwPDB-VP
Average B, all atoms (Å^2)	67.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 47.47 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 9.8796e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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	B	0.51	0/355	0.82	0/547
1	D	0.53	0/355	0.83	0/547
1	G	0.49	0/307	0.82	0/473
1	J	0.47	0/355	0.80	0/547
2	C	0.51	0/333	0.94	0/510
2	E	0.55	0/329	0.99	0/505
2	H	0.49	0/311	0.93	0/476
2	K	0.49	0/329	0.95	0/505
3	A	0.27	0/829	0.53	0/1114
3	F	0.27	0/847	0.51	0/1139
3	I	0.27	0/833	0.50	0/1120
3	L	0.28	0/804	0.50	0/1083
All	All	0.40	0/5987	0.72	0/8566

There are no bond length outliers.

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	B	315	0	168	4	0
1	D	315	0	168	2	0
1	G	273	0	146	3	0
1	J	315	0	168	3	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	C	300	0	171	4	0
2	E	296	0	169	2	0
2	H	280	0	159	4	0
2	K	296	0	169	2	0
3	A	809	0	809	11	0
3	F	827	0	814	3	0
3	I	813	0	807	9	0
3	L	784	0	765	7	0
4	A	12	0	0	0	0
4	C	1	0	0	0	0
4	D	2	0	0	0	0
4	E	2	0	0	0	0
4	F	9	0	0	0	0
4	I	2	0	0	0	0
4	J	1	0	0	0	0
4	K	1	0	0	0	0
4	L	2	0	0	0	0
All	All	5655	0	4513	47	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 5.

All (47) 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:B:15:DG:H1	2:C:2:DC:H42	1.47	0.62
3:A:410:ARG:HG2	3:A:411:LEU:HG	1.81	0.62
3:I:334:SER:O	3:I:379:HIS:NE2	2.32	0.58
2:H:8:DT:OP1	3:I:384:ASN:ND2	2.35	0.58
3:L:402:ILE:HG22	3:L:403:ILE:HG13	1.85	0.58
3:L:362:LYS:HB3	3:L:415:PHE:HB2	1.87	0.57
3:A:345:LEU:HD21	3:A:354:ILE:HG12	1.88	0.56
3:I:331:ILE:HG12	3:I:343:GLN:HB3	1.87	0.56
2:E:10:DC:H2"	2:E:11:DG:C8	2.41	0.55
1:G:12:DT:H2"	1:G:13:DG:C8	2.42	0.54
1:G:3:DA:H2"	1:G:4:DG:C8	2.43	0.54
3:A:348:SER:O	3:A:351:GLU:HG2	2.07	0.54
3:A:417:LYS:HB2	3:A:422:ILE:HG12	1.89	0.54
2:E:2:DC:H2'	2:E:3:DT:C6	2.44	0.53
2:C:10:DC:H2"	2:C:11:DG:C8	2.44	0.53
3:I:404:ARG:HH11	3:I:416:MET:HG3	1.73	0.53

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:J:12:DT:H2"	1:J:13:DG:C8	2.44	0.53
3:I:347:ASP:OD2	3:I:349:ARG:NH2	2.43	0.52
3:A:417:LYS:HB3	3:A:421:GLU:HB2	1.92	0.51
2:H:14:DT:H2"	3:L:383:THR:HG23	1.92	0.51
3:I:348:SER:O	3:I:351:GLU:HG2	2.12	0.50
1:D:12:DT:H2"	1:D:13:DG:C8	2.48	0.48
2:C:1:DT:H2'	2:C:2:DC:C6	2.48	0.47
3:A:342:TYR:OH	3:A:362:LYS:NZ	2.43	0.47
1:B:12:DT:H2"	1:B:13:DG:C8	2.50	0.47
3:I:399:LYS:HE3	3:I:399:LYS:HA	1.96	0.47
3:A:339:ASP:O	3:A:343:GLN:HG2	2.16	0.46
3:A:402:ILE:HG23	3:A:422:ILE:HD13	1.97	0.46
3:A:418:THR:O	3:A:422:ILE:HG13	2.16	0.46
1:D:1:DA:H2"	1:D:2:DA:OP1	2.15	0.46
1:J:8:DG:N7	3:L:392:ARG:NH1	2.64	0.45
3:I:408:GLY:H	3:L:378:ASN:HD22	1.62	0.45
3:F:348:SER:O	3:F:351:GLU:HG3	2.18	0.44
3:F:382:ARG:HH11	3:F:382:ARG:HG3	1.83	0.44
3:L:341:VAL:O	3:L:345:LEU:HG	2.19	0.43
2:H:10:DC:H2"	2:H:11:DG:C8	2.54	0.42
3:A:404:ARG:HG3	3:A:416:MET:HG3	2.01	0.42
2:H:6:DC:H2"	2:H:7:DA:C8	2.54	0.42
3:I:414:ARG:HH11	3:I:414:ARG:HG2	1.83	0.42
1:J:5:DC:H5"	3:L:405:LYS:HE3	2.01	0.42
3:F:330:ARG:HB3	3:F:333:ASP:HB2	2.01	0.42
2:K:10:DC:H2"	2:K:11:DG:C8	2.55	0.42
3:A:380:LYS:HE2	3:A:380:LYS:HB3	1.80	0.41
1:B:1:DA:H1'	1:B:2:DA:H5'	2.02	0.41
1:B:15:DG:C6	2:C:1:DT:H71	2.57	0.40
2:K:15:DT:OP2	2:K:15:DT:H2'	2.22	0.40
1:G:6:DC:H2'	1:G:7:DG:C8	2.57	0.40

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	
3	A	91/101 (90%)	90 (99%)	1 (1%)	0	100	100
3	F	95/101 (94%)	94 (99%)	1 (1%)	0	100	100
3	I	92/101 (91%)	91 (99%)	1 (1%)	0	100	100
3	L	90/101 (89%)	89 (99%)	1 (1%)	0	100	100
All	All	368/404 (91%)	364 (99%)	4 (1%)	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	
3	A	86/92 (94%)	84 (98%)	2 (2%)	50	67
3	F	87/92 (95%)	84 (97%)	3 (3%)	37	51
3	I	87/92 (95%)	83 (95%)	4 (5%)	27	38
3	L	83/92 (90%)	79 (95%)	4 (5%)	25	36
All	All	343/368 (93%)	330 (96%)	13 (4%)	33	47

All (13) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	333	ASP
3	A	423	MET
3	F	342	TYR
3	F	348	SER
3	F	360	GLU
3	I	342	TYR
3	I	357	GLU
3	I	405	LYS
3	I	410	ARG
3	L	346	SER

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	L	399	LYS
3	L	400	LEU
3	L	420	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (7) such sidechains are listed below:

Mol	Chain	Res	Type
3	A	343	GLN
3	A	409	GLN
3	F	384	ASN
3	I	352	ASN
3	L	352	ASN
3	L	378	ASN
3	L	401	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

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	B	15/15 (100%)	0.10	1 (6%) 17 23	64, 86, 123, 124	0
1	D	15/15 (100%)	-0.48	0 100 100	55, 59, 97, 102	0
1	G	13/15 (86%)	-0.24	0 100 100	73, 88, 104, 141	0
1	J	15/15 (100%)	-0.37	1 (6%) 17 23	54, 73, 89, 106	0
2	C	15/15 (100%)	-0.08	0 100 100	50, 79, 135, 165	0
2	E	15/15 (100%)	-0.50	0 100 100	42, 57, 83, 85	0
2	H	14/15 (93%)	0.07	0 100 100	53, 95, 129, 145	0
2	K	15/15 (100%)	-0.40	0 100 100	56, 82, 90, 90	0
3	A	93/101 (92%)	-0.03	1 (1%) 80 85	41, 50, 77, 93	0
3	F	97/101 (96%)	-0.01	0 100 100	36, 50, 79, 115	0
3	I	94/101 (93%)	0.29	3 (3%) 47 54	41, 57, 88, 105	0
3	L	92/101 (91%)	0.31	3 (3%) 46 53	43, 60, 90, 120	0
All	All	493/524 (94%)	0.05	9 (1%) 68 74	36, 58, 98, 165	0

All (9) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	L	422	ILE	3.4
3	L	334	SER	3.2
3	L	423	MET	2.9
1	B	1	DA	2.9
1	J	1	DA	2.9
3	I	331	ILE	2.9
3	I	335	ARG	2.8
3	I	333	ASP	2.7
3	A	331	ILE	2.6

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 monosaccharides in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.