W O R L D W I D E
wwPDB
P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 23, 2023 – 12:37 PM EDT

PDB ID	:	3DKU
Title	:	Crystal structure of Nudix hydrolase Orf153, ymfB, from Escherichia coli K-1
Authors	:	Hong, M.K.; Kim, J.K.; Jung, J.H.; Jung, J.W.; Choi, J.Y.; Kang, L.W.
Deposited on	:	2008-06-26
Resolution	:	2.69 Å(reported)

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

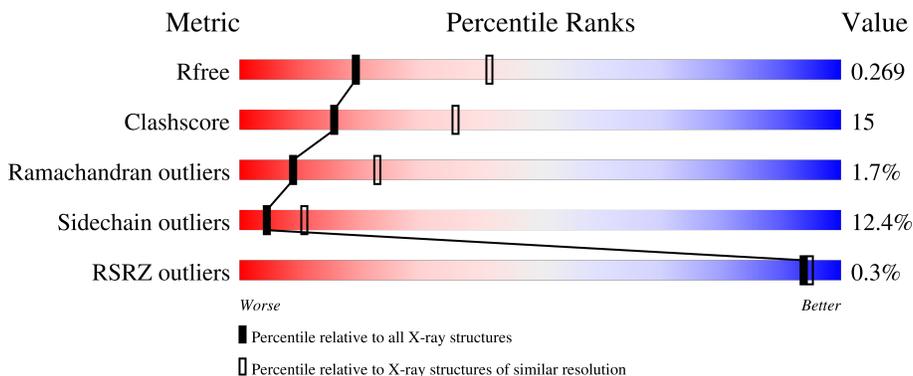|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 2.69 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	Percentile Ranks	Value
Rfree		0.269
Clashscore		15
Ramachandran outliers		1.7%
Sidechain outliers		12.4%
RSRZ outliers		0.3%

Worse   *Better*

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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	153	63% 31% 5%
1	B	153	58% 32% 5% · ·
1	C	153	70% 25% 5%
1	D	153	67% 24% 5% · ·
1	E	153	62% 30% 7% ·

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	153	65%　　27%　　5% ··
1	G	153	64%　　27%　　8% ·
1	H	153	.%　　59%　　29%　　6% · ·

2   Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 9800 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 Putative phosphohydrolase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	153	Total	C	N	O	S	0	0	0
			1227	786	209	224	8			
1	B	147	Total	C	N	O	S	0	0	0
			1181	754	202	217	8			
1	C	153	Total	C	N	O	S	0	0	0
			1227	786	209	224	8			
1	D	148	Total	C	N	O	S	0	0	0
			1192	763	203	218	8			
1	E	152	Total	C	N	O	S	0	0	0
			1219	780	208	223	8			
1	F	149	Total	C	N	O	S	0	0	0
			1199	767	204	220	8			
1	G	152	Total	C	N	O	S	0	0	0
			1219	780	208	223	8			
1	H	147	Total	C	N	O	S	0	0	0
			1181	754	202	217	8			

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	18	Total	O	0	0
			18	18		
2	B	19	Total	O	0	0
			19	19		
2	C	27	Total	O	0	0
			27	27		
2	D	20	Total	O	0	0
			20	20		
2	E	20	Total	O	0	0
			20	20		
2	F	22	Total	O	0	0
			22	22		

Continued on next page...

Continued from previous page...

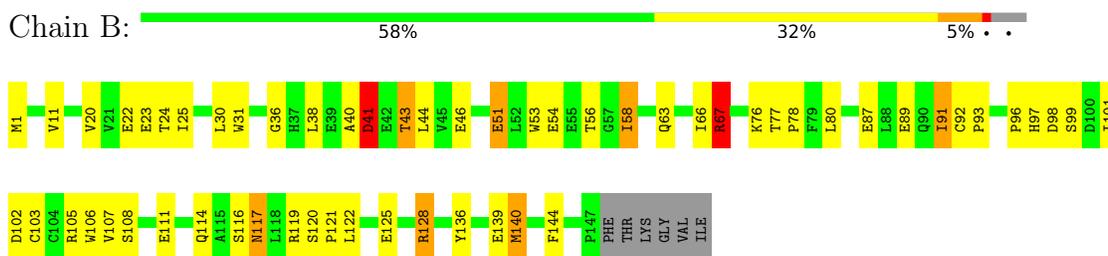| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	G	9	Total 9	O 9	0	0
2	H	20	Total 20	O 20	0	0

3 Residue-property plots ⓘ

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.
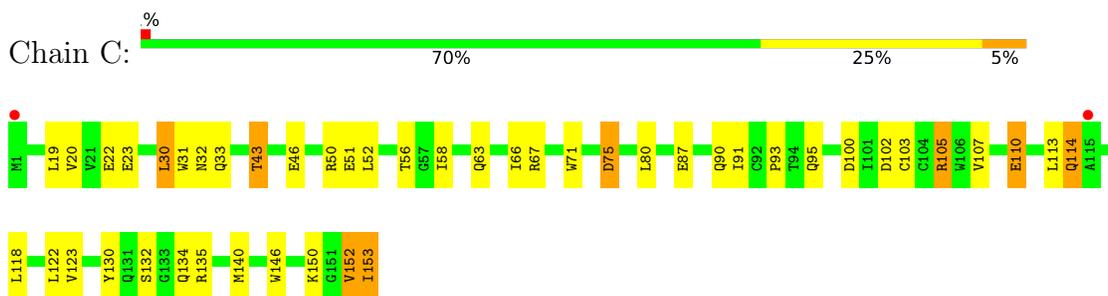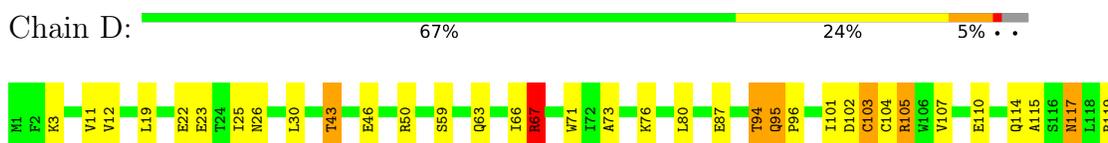
• Molecule 1: Putative phosphohydrolase

• Molecule 1: Putative phosphohydrolase
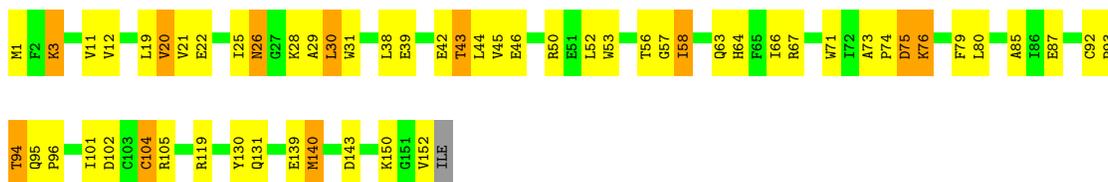

• Molecule 1: Putative phosphohydrolase

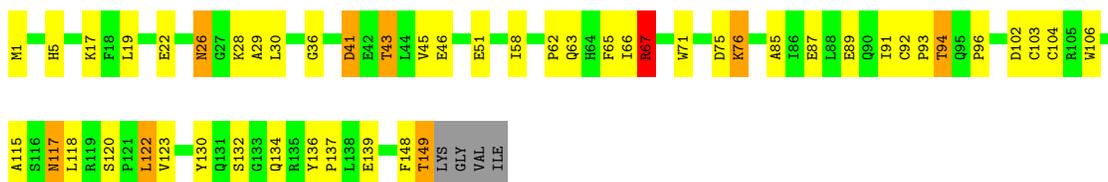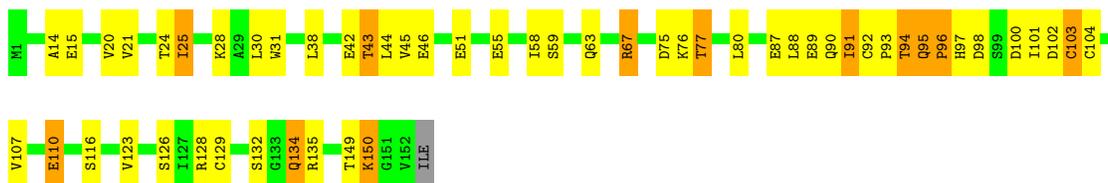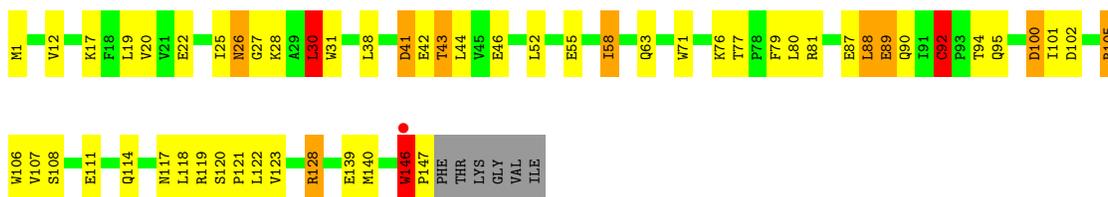• Molecule 1: Putative phosphohydrolase

E125 S126 I127 Y130 Q134 R135 Y136 P137 M140 W146 P147 F148 THR LYS GLY VAL ILE

• Molecule 1: Putative phosphohydrolase

Chain E:　　　　62%　　　　30%　　　　7% •

M1 F2 K3 V11 V12 L19 V20 V21 E22 I25 N26 G27 K28 A29 L30 W31 L38 E39 E42 T43 L44 V45 E46 R50 E51 L52 W53 T56 G57 I58 Q63 H64 F65 I66 R67 W71 I72 A73 P74 D75 K76 F79 L80 A85 I86 E87 C92 P93

T94 Q95 P96 I101 D102 C103 C104 R105 R119 Y130 Q131 E139 M140 D143 K150 G151 V152 ILE

• Molecule 1: Putative phosphohydrolase

Chain F:　　　　65%　　　　27%　　　　5% • •

M1 H5 K17 F18 L19 E22 N26 G27 K28 A29 L30 G36 D41 E42 T43 L44 V45 E46 E51 I58 P62 Q63 H64 F65 I66 R67 W71 D75 K76 A85 I86 E87 L88 E89 Q90 I91 C92 P93 T94 Q95 P96 D102 C103 C104 R105 W106

A115 S116 N117 L118 R119 S120 P121 L122 V123 Y130 Q131 S132 G133 Q134 R135 Y136 P137 L138 E139 F148 T149 LYS GLY VAL ILE

• Molecule 1: Putative phosphohydrolase

Chain G:　　　　64%　　　　27%　　　　8% •

M1 A14 E15 V20 V21 T24 I25 K28 A29 L30 W31 L38 E42 T43 L44 V45 E46 E51 E55 I58 S59 Q63 R67 D75 K76 T77 L80 E87 L88 E89 Q90 I91 C92 P93 T94 Q95 P96 H97 D98 S99 D100 I101 D102 C103 C104

V107 E110 S116 V123 S126 I127 R128 C129 S132 G133 Q134 R135 T149 K150 G151 V152 ILE

• Molecule 1: Putative phosphohydrolase

%

Chain H:　　　　59%　　　　29%　　　　6% • •

M1 V12 K17 F18 L19 V20 V21 E22 I25 N26 G27 K28 A29 L30 W31 L38 D41 E42 T43 L44 V45 E46 L52 E55 I58 Q63 W71 K76 T77 P78 F79 L80 R81 E87 L88 E89 Q90 I91 C92 P93 T94 Q95 D100 I101 D102 R105

W106 V107 S108 E111 Q114 N117 L118 R119 S120 P121 L122 V123 R128 E139 M140 W146 P147 PHE THR LYS GLY VAL ILE

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	68.96Å 70.52Å 145.03Å 90.00° 103.30° 90.00°	Depositor
Resolution (Å)	33.56 – 2.69 33.56 – 2.69	Depositor EDS
% Data completeness (in resolution range)	98.3 (33.56-2.69) 98.3 (33.56-2.69)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.60 (at 2.68Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.193 , 0.270 0.194 , 0.269	Depositor DCC
R_{free} test set	1865 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å^2)	33.9	Xtriage
Anisotropy	0.130	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 35.8	EDS
L-test for twinning [2]	$< |L| > = 0.49, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.019 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	9800	wwPDB-VP
Average B, all atoms (Å^2)	28.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.40% 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	0.85	0/1262	0.84	0/1719
1	B	0.90	2/1215 (0.2%)	0.89	1/1656 (0.1%)
1	C	0.83	0/1262	0.85	2/1719 (0.1%)
1	D	0.86	0/1227	0.93	3/1672 (0.2%)
1	E	0.92	1/1254 (0.1%)	0.84	1/1708 (0.1%)
1	F	0.94	2/1234 (0.2%)	0.83	0/1682
1	G	0.89	1/1254 (0.1%)	0.88	2/1708 (0.1%)
1	H	0.92	1/1215 (0.1%)	0.92	4/1656 (0.2%)
All	All	0.89	7/9923 (0.1%)	0.87	13/13520 (0.1%)

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	H	0	1

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	H	92	CYS	CB-SG	-11.05	1.63	1.82
1	G	129	CYS	CB-SG	-6.64	1.71	1.82
1	E	139	GLU	CG-CD	6.52	1.61	1.51
1	B	139	GLU	CG-CD	5.61	1.60	1.51
1	B	51	GLU	CG-CD	5.52	1.60	1.51

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	H	146	TRP	C-N-CD	-7.83	103.37	120.60
1	D	146	TRP	C-N-CD	-7.48	104.15	120.60
1	H	146	TRP	C-N-CA	6.32	148.54	122.00

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	H	30	LEU	CA-CB-CG	6.19	129.53	115.30
1	G	67	ARG	NE-CZ-NH1	-6.06	117.27	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	H	146	TRP	Peptide

5.2 Too-close contacts (i)

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	1227	0	1190	45	0
1	B	1181	0	1138	38	0
1	C	1227	0	1190	22	0
1	D	1192	0	1147	30	0
1	E	1219	0	1179	43	0
1	F	1199	0	1154	38	0
1	G	1219	0	1179	36	0
1	H	1181	0	1138	42	0
2	A	18	0	0	0	0
2	B	19	0	0	2	0
2	C	27	0	0	1	0
2	D	20	0	0	0	0
2	E	20	0	0	3	0
2	F	22	0	0	1	0
2	G	9	0	0	0	0
2	H	20	0	0	2	0
All	All	9800	0	9315	278	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 15.

The worst 5 of 278 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:25:ILE:HG23	1:A:26:ASN:H	1.11	1.06
1:B:36:GLY:HA3	1:B:51:GLU:HG3	1.38	1.06
1:G:58:ILE:HD12	1:G:92:CYS:SG	2.02	0.99
1:H:43:THR:HG22	1:H:46:GLU:H	1.25	0.98
1:F:63:GLN:NE2	1:F:87:GLU:H	1.65	0.95

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	151/153 (99%)	139 (92%)	8 (5%)	4 (3%)	5	13
1	B	145/153 (95%)	135 (93%)	7 (5%)	3 (2%)	7	18
1	C	151/153 (99%)	137 (91%)	13 (9%)	1 (1%)	22	46
1	D	146/153 (95%)	132 (90%)	11 (8%)	3 (2%)	7	18
1	E	150/153 (98%)	134 (89%)	14 (9%)	2 (1%)	12	30
1	F	147/153 (96%)	134 (91%)	11 (8%)	2 (1%)	11	28
1	G	150/153 (98%)	134 (89%)	13 (9%)	3 (2%)	7	19
1	H	145/153 (95%)	131 (90%)	12 (8%)	2 (1%)	11	28
All	All	1185/1224 (97%)	1076 (91%)	89 (8%)	20 (2%)	9	23

5 of 20 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	G	100	ASP
1	H	27	GLY
1	A	57	GLY
1	C	75	ASP
1	F	41	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	132/132 (100%)	119 (90%)	13 (10%)	8	18
1	B	127/132 (96%)	108 (85%)	19 (15%)	3	7
1	C	132/132 (100%)	116 (88%)	16 (12%)	5	11
1	D	128/132 (97%)	114 (89%)	14 (11%)	6	14
1	E	131/132 (99%)	119 (91%)	12 (9%)	9	21
1	F	129/132 (98%)	117 (91%)	12 (9%)	9	21
1	G	131/132 (99%)	110 (84%)	21 (16%)	2	6
1	H	127/132 (96%)	105 (83%)	22 (17%)	2	5
All	All	1037/1056 (98%)	908 (88%)	129 (12%)	4	11

5 of 129 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	H	58	ILE
1	H	89	GLU
1	D	43	THR
1	C	153	ILE
1	H	94	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 31 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	134	GLN
1	G	95	GLN
1	E	61	GLN
1	H	63	GLN
1	F	117	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, 95th 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(Å²)	Q<0.9
1	A	153/153 (100%)	-0.41	1 (0%)	87	89	15, 29, 44, 51	2 (1%)
1	B	147/153 (96%)	-0.43	0	100	100	15, 26, 39, 56	0
1	C	153/153 (100%)	-0.31	2 (1%)	77	78	13, 27, 39, 44	0
1	D	148/153 (96%)	-0.68	0	100	100	13, 24, 35, 44	2 (1%)
1	E	152/153 (99%)	-0.44	0	100	100	15, 26, 40, 53	0
1	F	149/153 (97%)	-0.49	0	100	100	11, 27, 42, 53	1 (0%)
1	G	152/153 (99%)	-0.45	0	100	100	15, 29, 45, 55	0
1	H	147/153 (96%)	-0.49	1 (0%)	87	89	15, 27, 45, 62	0
All	All	1201/1224 (98%)	-0.46	4 (0%)	94	95	11, 27, 43, 62	5 (0%)

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	1	MET	3.2
1	H	146	TRP	3.1
1	C	115	ALA	2.2
1	A	139	GLU	2.0

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 ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.