



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 10:18 PM GMT

PDB ID : 4X8F  
Title : Vibrio cholerae O395 Ribokinase in apo form  
Authors : Paul, R.; Patra, M.D.; Sen, U.  
Deposited on : 2014-12-10  
Resolution : 3.40 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](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity : 4.02b-467  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk26865

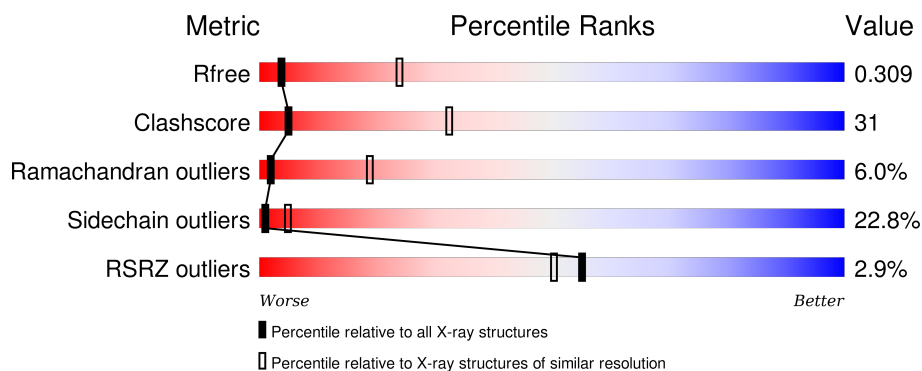
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.40 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))
$R_{free}$	91344	1476 (3.50-3.30)
Clashscore	102246	1611 (3.50-3.30)
Ramachandran outliers	100387	1571 (3.50-3.30)
Sidechain outliers	100360	1571 (3.50-3.30)
RSRZ outliers	91569	1485 (3.50-3.30)

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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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  $\leq 5\%$ . 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	309	<div> <div>6%</div> <div>43% 42% 14%</div> </div>
1	B	309	<div> <div>4%</div> <div>36% 45% 17%</div> </div>
1	C	309	<div> <div>3%</div> <div>40% 44% 14%</div> </div>
1	D	309	<div> <div>%</div> <div>40% 44% 14%</div> </div>
1	E	309	<div> <div>3%</div> <div>47% 42% 10%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	F	309	
1	G	309	
1	H	309	
1	I	309	
1	J	309	
1	K	309	
1	L	309	
1	M	309	
1	N	309	
1	O	309	
1	P	309	

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 36000 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 Ribokinase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	306	Total	C	N	O	S	0	0	0
			2250	1408	391	441	10			
1	B	306	Total	C	N	O	S	0	0	0
			2250	1408	391	441	10			
1	C	306	Total	C	N	O	S	0	0	0
			2250	1408	391	441	10			
1	D	306	Total	C	N	O	S	0	0	0
			2250	1408	391	441	10			
1	E	306	Total	C	N	O	S	0	0	0
			2250	1408	391	441	10			
1	F	306	Total	C	N	O	S	0	0	0
			2250	1408	391	441	10			
1	G	306	Total	C	N	O	S	0	0	0
			2250	1408	391	441	10			
1	H	306	Total	C	N	O	S	0	0	0
			2250	1408	391	441	10			
1	I	306	Total	C	N	O	S	0	0	0
			2250	1408	391	441	10			
1	J	306	Total	C	N	O	S	0	0	0
			2250	1408	391	441	10			
1	K	306	Total	C	N	O	S	0	0	0
			2250	1408	391	441	10			
1	L	306	Total	C	N	O	S	0	0	0
			2250	1408	391	441	10			
1	M	306	Total	C	N	O	S	0	0	0
			2250	1408	391	441	10			
1	N	306	Total	C	N	O	S	0	0	0
			2250	1408	391	441	10			
1	O	306	Total	C	N	O	S	0	0	0
			2250	1408	391	441	10			
1	P	306	Total	C	N	O	S	0	0	0
			2250	1408	391	441	10			

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	expression tag	UNP A5F1B7
A	-1	SER	-	expression tag	UNP A5F1B7
A	0	HIS	-	expression tag	UNP A5F1B7
B	-2	GLY	-	expression tag	UNP A5F1B7
B	-1	SER	-	expression tag	UNP A5F1B7
B	0	HIS	-	expression tag	UNP A5F1B7
C	-2	GLY	-	expression tag	UNP A5F1B7
C	-1	SER	-	expression tag	UNP A5F1B7
C	0	HIS	-	expression tag	UNP A5F1B7
D	-2	GLY	-	expression tag	UNP A5F1B7
D	-1	SER	-	expression tag	UNP A5F1B7
D	0	HIS	-	expression tag	UNP A5F1B7
E	-2	GLY	-	expression tag	UNP A5F1B7
E	-1	SER	-	expression tag	UNP A5F1B7
E	0	HIS	-	expression tag	UNP A5F1B7
F	-2	GLY	-	expression tag	UNP A5F1B7
F	-1	SER	-	expression tag	UNP A5F1B7
F	0	HIS	-	expression tag	UNP A5F1B7
G	-2	GLY	-	expression tag	UNP A5F1B7
G	-1	SER	-	expression tag	UNP A5F1B7
G	0	HIS	-	expression tag	UNP A5F1B7
H	-2	GLY	-	expression tag	UNP A5F1B7
H	-1	SER	-	expression tag	UNP A5F1B7
H	0	HIS	-	expression tag	UNP A5F1B7
I	-2	GLY	-	expression tag	UNP A5F1B7
I	-1	SER	-	expression tag	UNP A5F1B7
I	0	HIS	-	expression tag	UNP A5F1B7
J	-2	GLY	-	expression tag	UNP A5F1B7
J	-1	SER	-	expression tag	UNP A5F1B7
J	0	HIS	-	expression tag	UNP A5F1B7
K	-2	GLY	-	expression tag	UNP A5F1B7
K	-1	SER	-	expression tag	UNP A5F1B7
K	0	HIS	-	expression tag	UNP A5F1B7
L	-2	GLY	-	expression tag	UNP A5F1B7
L	-1	SER	-	expression tag	UNP A5F1B7
L	0	HIS	-	expression tag	UNP A5F1B7
M	-2	GLY	-	expression tag	UNP A5F1B7
M	-1	SER	-	expression tag	UNP A5F1B7
M	0	HIS	-	expression tag	UNP A5F1B7
N	-2	GLY	-	expression tag	UNP A5F1B7
N	-1	SER	-	expression tag	UNP A5F1B7
N	0	HIS	-	expression tag	UNP A5F1B7

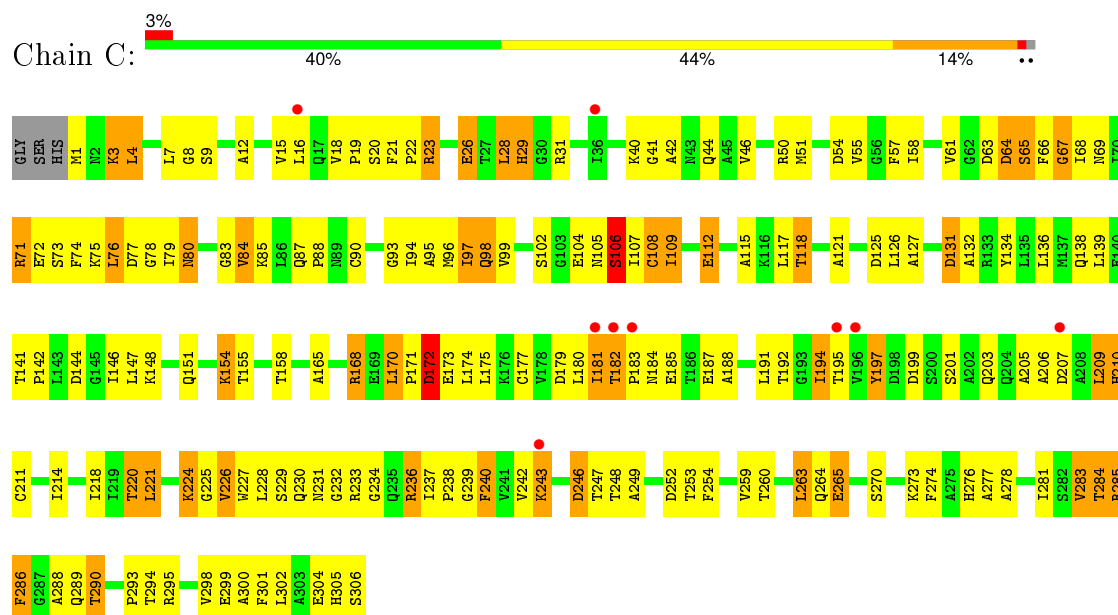
*Continued on next page...*

*Continued from previous page...*

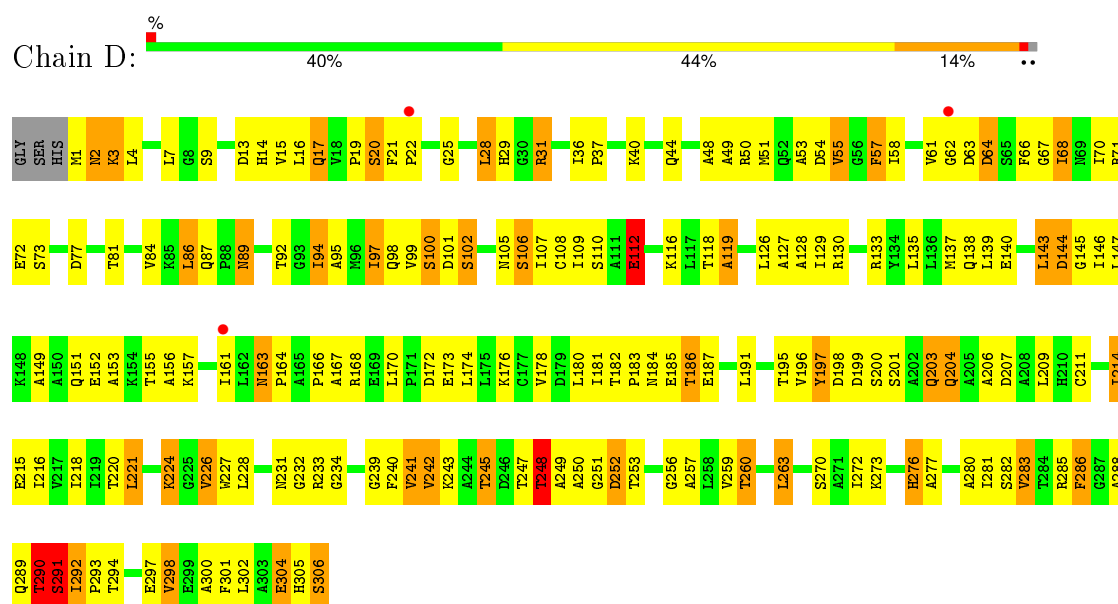
Chain	Residue	Modelled	Actual	Comment	Reference
O	-2	GLY	-	expression tag	UNP A5F1B7
O	-1	SER	-	expression tag	UNP A5F1B7
O	0	HIS	-	expression tag	UNP A5F1B7
P	-2	GLY	-	expression tag	UNP A5F1B7
P	-1	SER	-	expression tag	UNP A5F1B7
P	0	HIS	-	expression tag	UNP A5F1B7



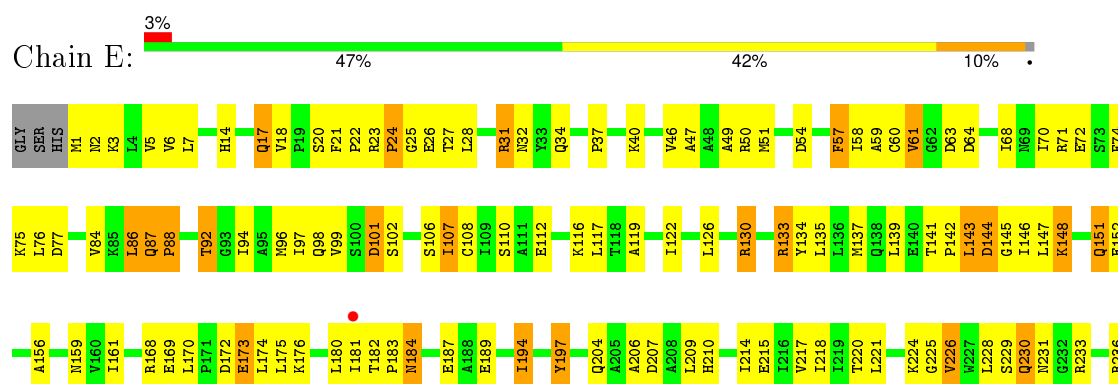
- Molecule 1: Ribokinase



- Molecule 1: Ribokinase



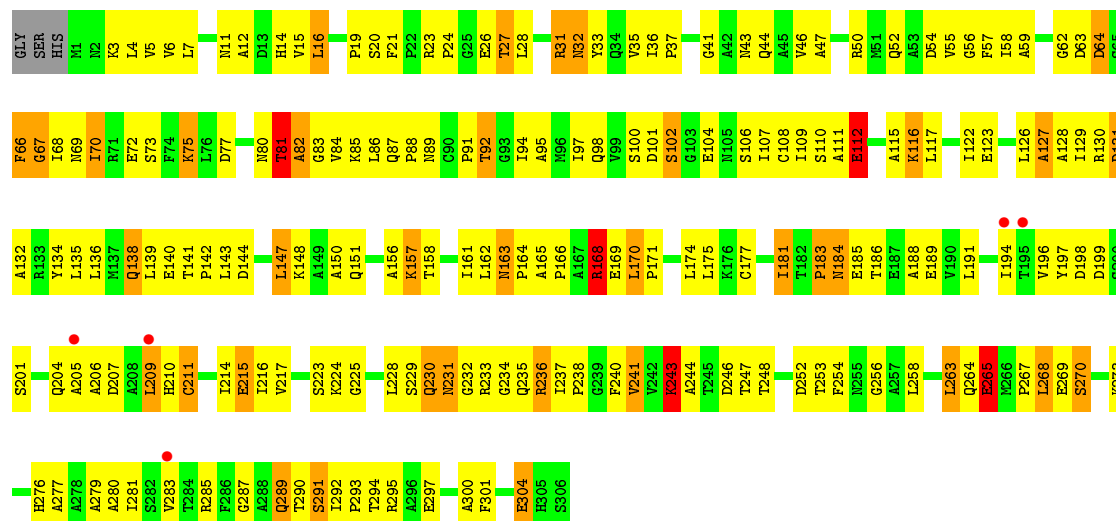
- Molecule 1: Ribokinase



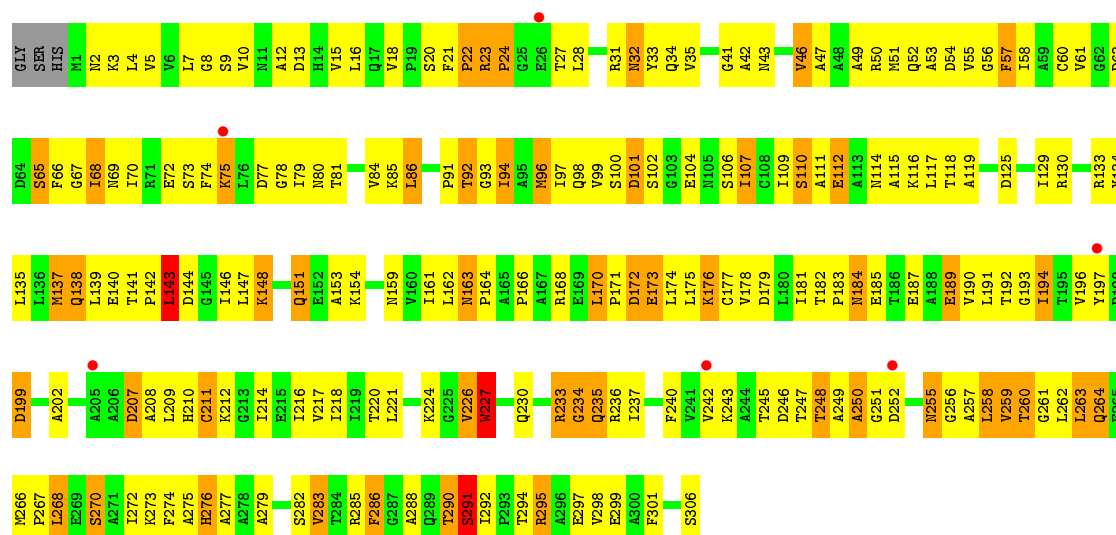




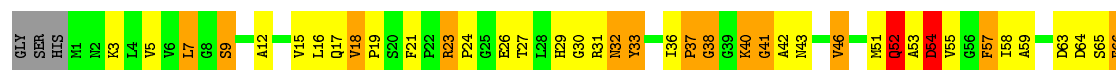
• Molecule 1: Ribokinase

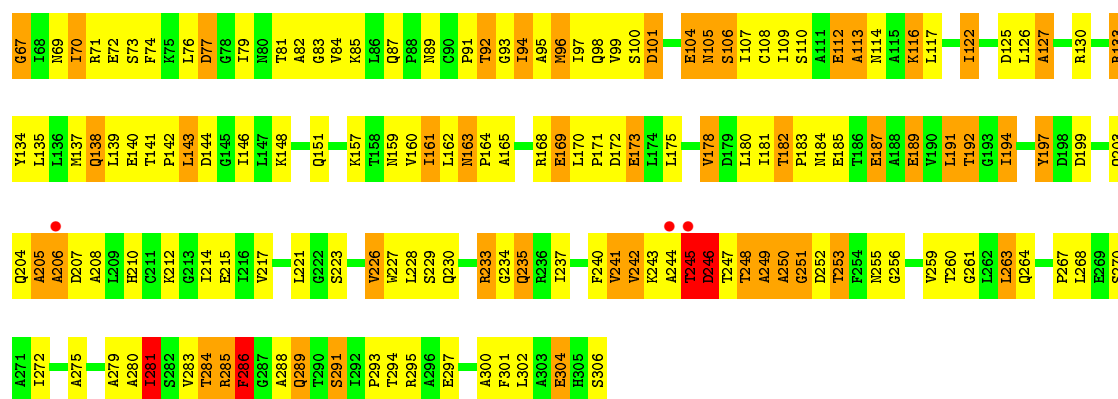


• Molecule 1: Ribokinase

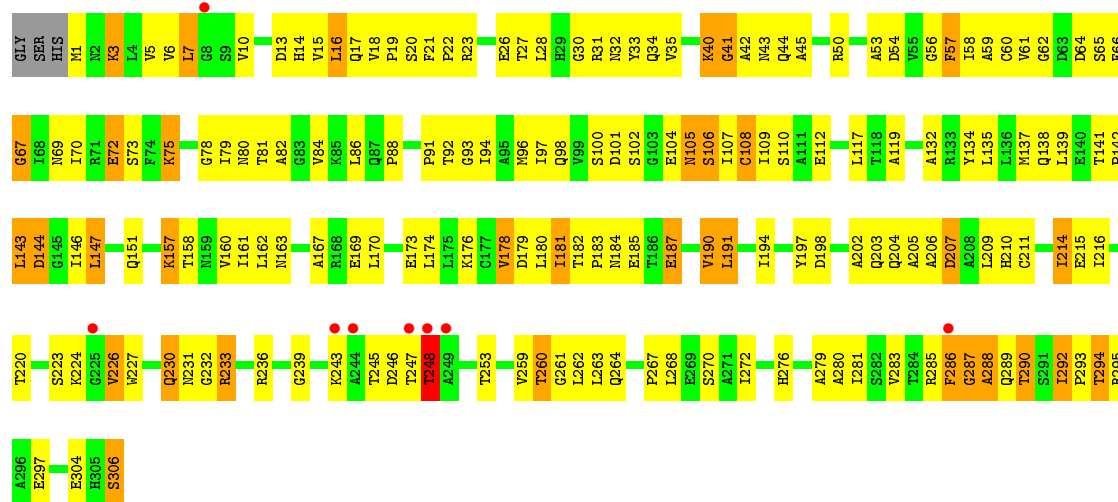


• Molecule 1: Ribokinase

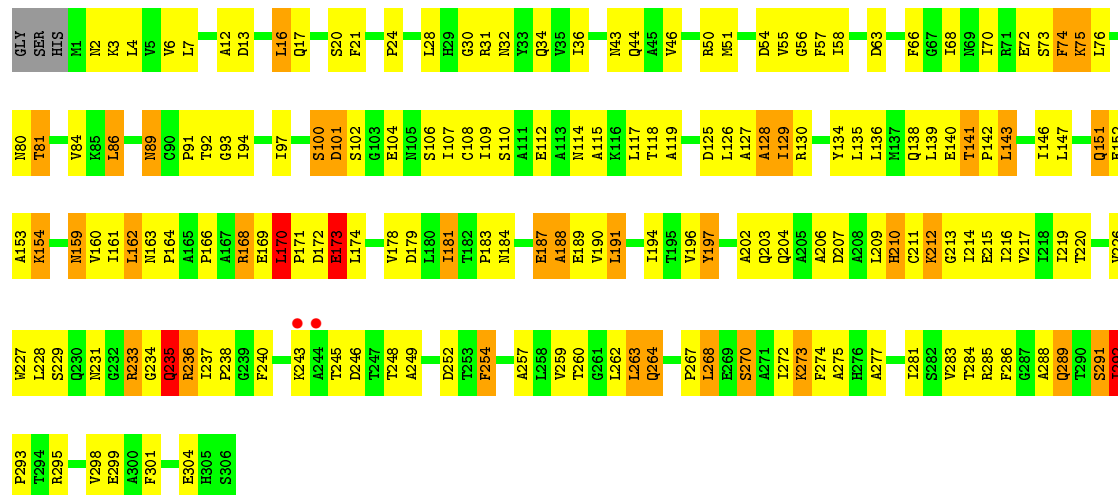




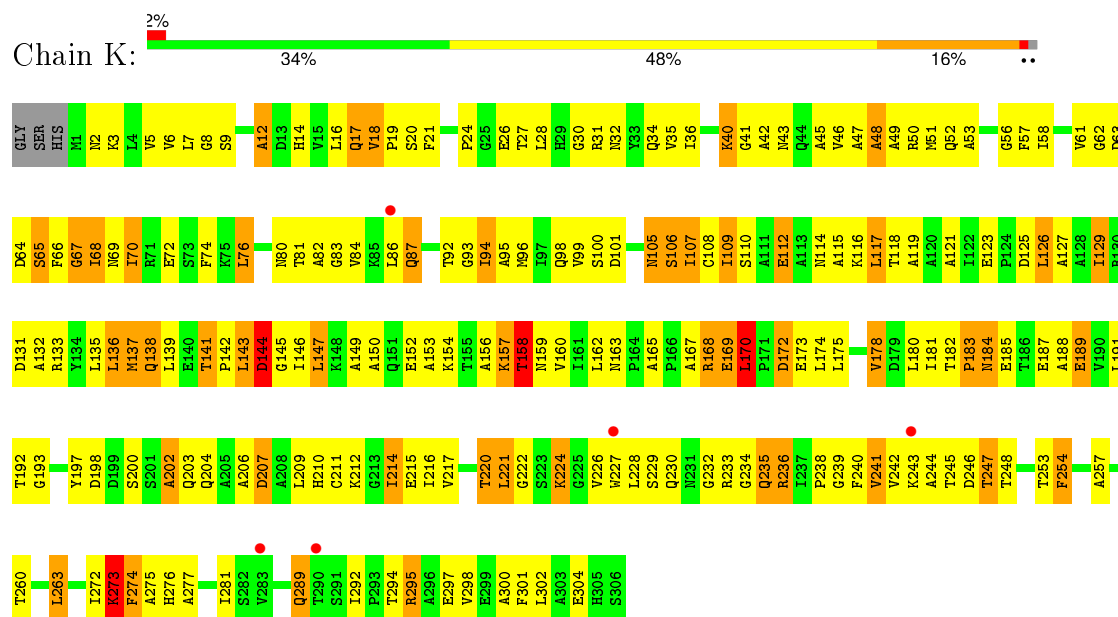
### • Molecule 1: Ribokinase



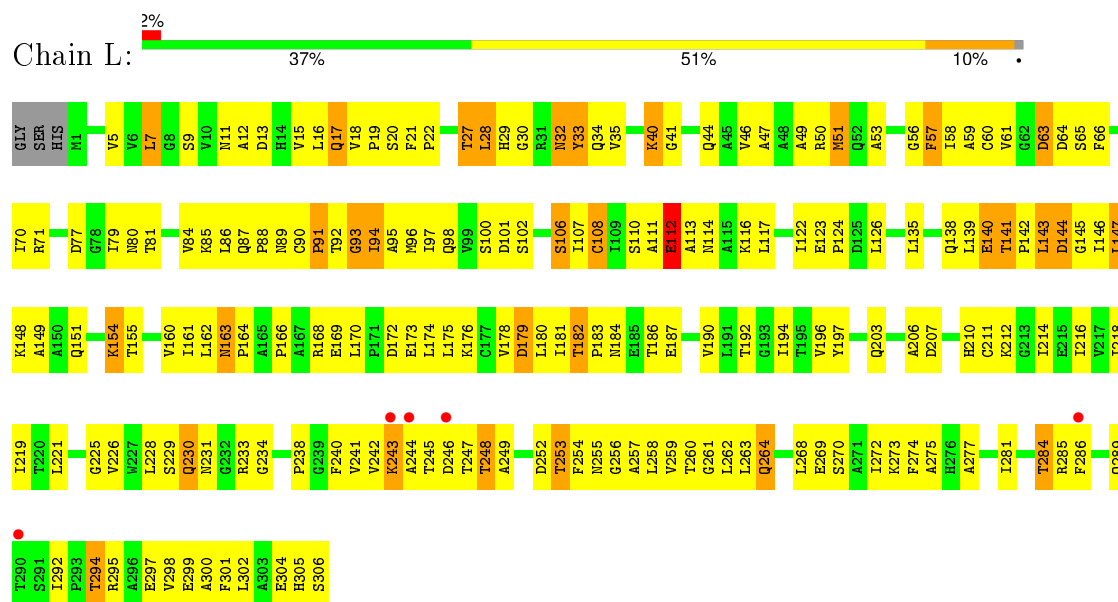
### • Molecule 1: Ribokinase



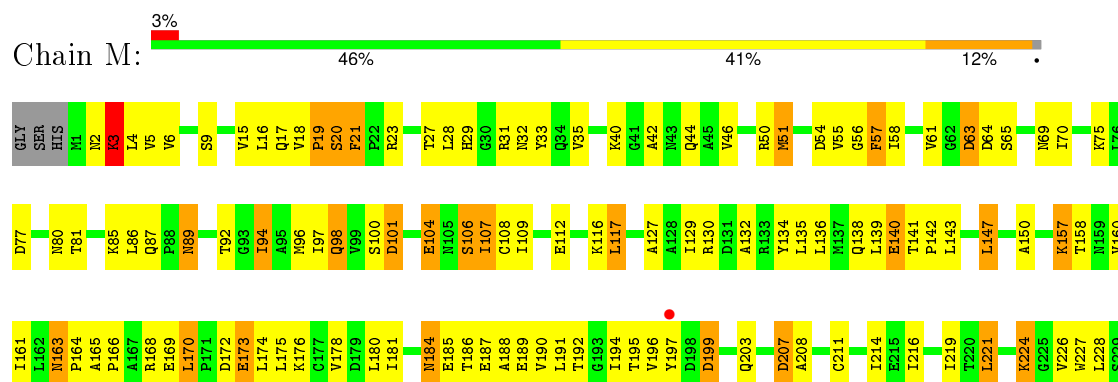
- Molecule 1: Ribokinase

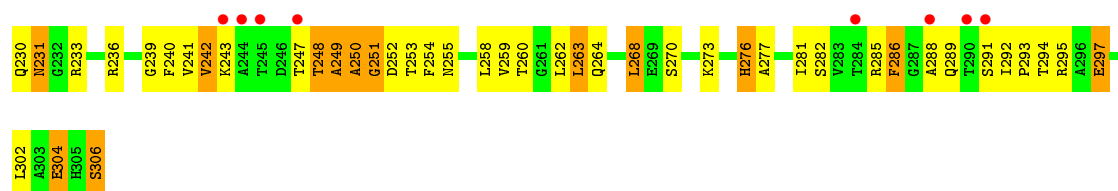


- Molecule 1: Ribokinase

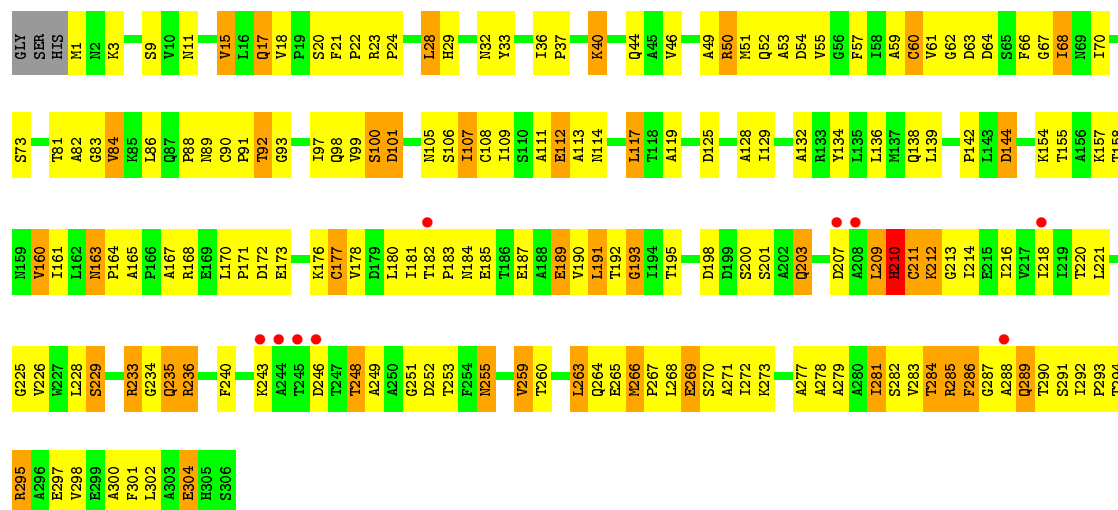


- Molecule 1: Ribokinase

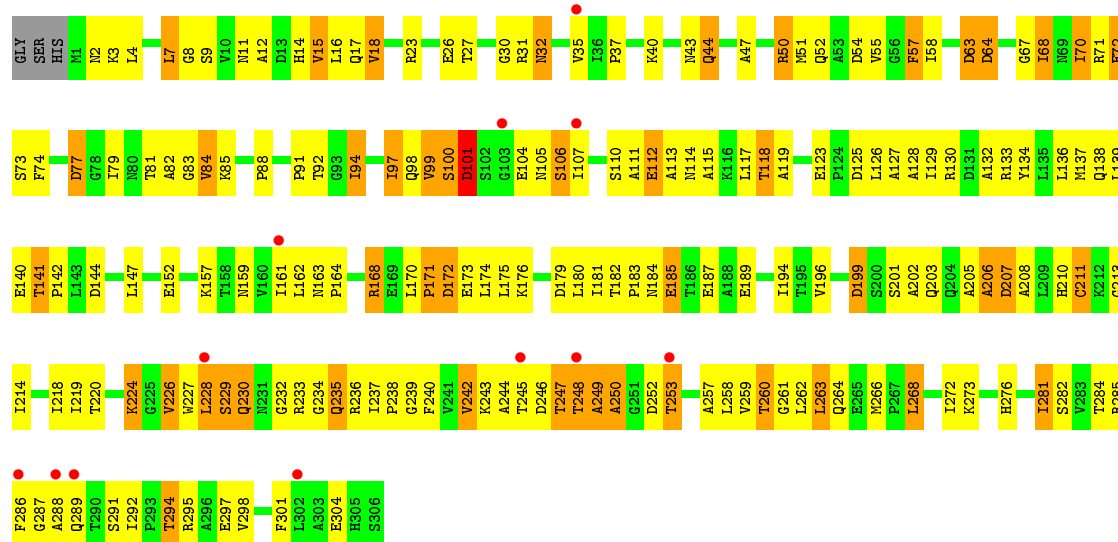




## ● Molecule 1: Ribokinase

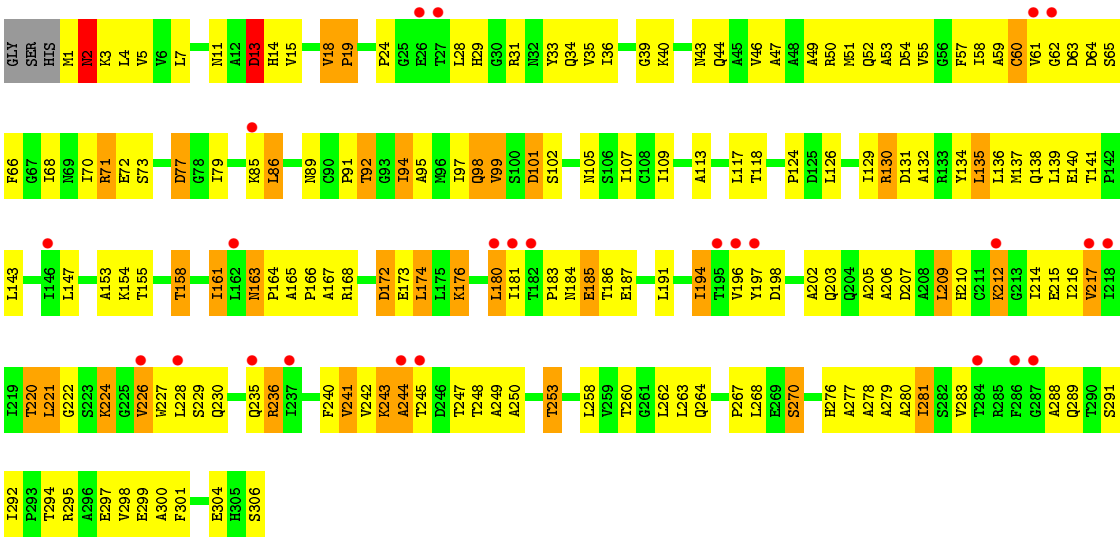


## ● Molecule 1: Ribokinase



## ● Molecule 1: Ribokinase





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	129.22Å 130.85Å 145.69Å 110.52° 90.00° 119.59°	Depositor
Resolution (Å)	43.57 – 3.40 49.34 – 3.40	Depositor EDS
% Data completeness (in resolution range)	92.8 (43.57-3.40) 79.0 (49.34-3.40)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.15	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.13 (at 3.40Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, $R_{free}$	0.239 , 0.308 0.244 , 0.309	Depositor DCC
$R_{free}$ test set	4832 reflections (5.24%)	DCC
Wilson B-factor (Å <sup>2</sup> )	63.2	Xtriage
Anisotropy	0.563	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 52.8	EDS
Estimated twinning fraction	0.277 for h,-h-k,-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.44$ , $\langle L^2 \rangle = 0.26$	Xtriage
Outliers	0 of 97031 reflections	Xtriage
$F_o, F_c$ correlation	0.88	EDS
Total number of atoms	36000	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	88.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.86% of the height of the origin peak. No significant pseudotranslation is detected.*

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

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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	0.48	0/2283	0.81	1/3105 (0.0%)
1	B	0.48	0/2283	0.81	0/3105
1	C	0.52	0/2283	0.86	1/3105 (0.0%)
1	D	0.56	0/2283	0.91	1/3105 (0.0%)
1	E	0.62	0/2283	0.91	0/3105
1	F	0.55	0/2283	0.83	1/3105 (0.0%)
1	G	0.64	1/2283 (0.0%)	0.93	1/3105 (0.0%)
1	H	0.64	0/2283	0.95	1/3105 (0.0%)
1	I	0.56	0/2283	0.83	0/3105
1	J	0.55	0/2283	0.85	1/3105 (0.0%)
1	K	0.57	0/2283	0.93	3/3105 (0.1%)
1	L	0.62	0/2283	0.85	0/3105
1	M	0.56	0/2283	0.84	0/3105
1	N	0.51	0/2283	0.80	0/3105
1	O	0.47	0/2283	0.80	1/3105 (0.0%)
1	P	0.43	0/2283	0.74	0/3105
All	All	0.55	1/36528 (0.0%)	0.86	11/49680 (0.0%)

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	D	0	1
1	G	0	1
1	J	0	1
1	K	0	2
1	N	0	1
All	All	0	6

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	G	227	TRP	CB-CG	-5.13	1.41	1.50

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	G	143	LEU	CA-CB-CG	8.59	135.05	115.30
1	A	93	GLY	N-CA-C	-7.56	94.21	113.10
1	J	170	LEU	CA-CB-CG	7.36	132.22	115.30
1	F	168	ARG	NE-CZ-NH1	-6.72	116.94	120.30
1	D	86	LEU	CA-CB-CG	6.60	130.47	115.30

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	290	THR	Peptide
1	G	100	SER	Peptide
1	J	100	SER	Peptide
1	K	156	ALA	Peptide
1	K	157	LYS	Peptide

## 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	2250	0	2273	147	0
1	B	2250	0	2273	184	0
1	C	2250	0	2273	125	0
1	D	2250	0	2273	165	0
1	E	2250	0	2273	133	0
1	F	2250	0	2273	145	0
1	G	2250	0	2273	165	0
1	H	2250	0	2273	203	0
1	I	2250	0	2273	123	0
1	J	2250	0	2273	123	0
1	K	2250	0	2273	170	0
1	L	2250	0	2273	167	0
1	M	2250	0	2273	119	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	N	2250	0	2273	126	0
1	O	2250	0	2273	134	0
1	P	2250	0	2273	132	0
All	All	36000	0	36368	2252	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 31.

The worst 5 of 2252 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:H:31:ARG:NH2	1:H:101:ASP:OD1	1.89	1.06
1:K:94:ILE:HD12	1:K:95:ALA:H	1.18	1.04
1:L:51:MET:HB2	1:L:260:THR:HG21	1.43	1.00
1:H:243:LYS:HG3	1:K:101:ASP:HA	1.42	1.00
1:E:87:GLN:HG2	1:E:116:LYS:HG3	1.45	0.99

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	304/309 (98%)	242 (80%)	46 (15%)	16 (5%)	2	22
1	B	304/309 (98%)	228 (75%)	50 (16%)	26 (9%)	1	10
1	C	304/309 (98%)	236 (78%)	46 (15%)	22 (7%)	1	14
1	D	304/309 (98%)	251 (83%)	33 (11%)	20 (7%)	1	16
1	E	304/309 (98%)	261 (86%)	34 (11%)	9 (3%)	5	39
1	F	304/309 (98%)	239 (79%)	46 (15%)	19 (6%)	2	18
1	G	304/309 (98%)	236 (78%)	50 (16%)	18 (6%)	2	19

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	H	304/309 (98%)	238 (78%)	38 (12%)	28 (9%)	1	9
1	I	304/309 (98%)	264 (87%)	27 (9%)	13 (4%)	3	29
1	J	304/309 (98%)	244 (80%)	46 (15%)	14 (5%)	3	26
1	K	304/309 (98%)	238 (78%)	39 (13%)	27 (9%)	1	9
1	L	304/309 (98%)	259 (85%)	30 (10%)	15 (5%)	3	25
1	M	304/309 (98%)	253 (83%)	31 (10%)	20 (7%)	1	16
1	N	304/309 (98%)	262 (86%)	27 (9%)	15 (5%)	3	25
1	O	304/309 (98%)	245 (81%)	40 (13%)	19 (6%)	2	18
1	P	304/309 (98%)	259 (85%)	36 (12%)	9 (3%)	5	39
All	All	4864/4944 (98%)	3955 (81%)	619 (13%)	290 (6%)	2	19

5 of 290 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	64	ASP
1	A	67	GLY
1	A	82	ALA
1	A	172	ASP
1	A	196	VAL

### 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	237/239 (99%)	186 (78%)	51 (22%)	1	6
1	B	237/239 (99%)	171 (72%)	66 (28%)	0	3
1	C	237/239 (99%)	180 (76%)	57 (24%)	1	4
1	D	237/239 (99%)	185 (78%)	52 (22%)	1	6
1	E	237/239 (99%)	194 (82%)	43 (18%)	2	11
1	F	237/239 (99%)	181 (76%)	56 (24%)	1	4
1	G	237/239 (99%)	160 (68%)	77 (32%)	0	2

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	H	237/239 (99%)	178 (75%)	59 (25%)	1	4
1	I	237/239 (99%)	192 (81%)	45 (19%)	2	9
1	J	237/239 (99%)	185 (78%)	52 (22%)	1	6
1	K	237/239 (99%)	183 (77%)	54 (23%)	1	5
1	L	237/239 (99%)	197 (83%)	40 (17%)	2	14
1	M	237/239 (99%)	188 (79%)	49 (21%)	1	7
1	N	237/239 (99%)	186 (78%)	51 (22%)	1	6
1	O	237/239 (99%)	181 (76%)	56 (24%)	1	4
1	P	237/239 (99%)	180 (76%)	57 (24%)	1	4
All	All	3792/3824 (99%)	2927 (77%)	865 (23%)	1	5

5 of 865 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	268	LEU
1	I	178	VAL
1	O	248	THR
1	H	7	LEU
1	H	191	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 82 such sidechains are listed below:

Mol	Chain	Res	Type
1	I	43	ASN
1	J	163	ASN
1	O	289	GLN
1	I	98	GLN
1	J	17	GLN

### 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 [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

There are no ligands in this entry.

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

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	306/309 (99%)	0.27	19 (6%)	24	22	61, 108, 179, 294	0
1	B	306/309 (99%)	0.21	12 (3%)	43	38	67, 98, 149, 239	0
1	C	306/309 (99%)	0.08	9 (2%)	55	50	42, 83, 141, 201	0
1	D	306/309 (99%)	-0.03	3 (0%)	84	79	34, 76, 134, 179	0
1	E	306/309 (99%)	-0.03	9 (2%)	55	50	38, 62, 112, 182	0
1	F	306/309 (99%)	-0.02	5 (1%)	74	69	46, 81, 140, 173	0
1	G	306/309 (99%)	0.00	6 (1%)	68	62	40, 79, 131, 162	0
1	H	306/309 (99%)	-0.04	3 (0%)	84	79	36, 69, 125, 214	0
1	I	306/309 (99%)	0.03	8 (2%)	59	54	43, 76, 125, 199	0
1	J	306/309 (99%)	-0.00	2 (0%)	89	85	43, 80, 130, 202	0
1	K	306/309 (99%)	-0.06	5 (1%)	74	69	47, 83, 140, 172	0
1	L	306/309 (99%)	-0.02	5 (1%)	74	69	36, 63, 112, 186	0
1	M	306/309 (99%)	0.11	9 (2%)	55	50	42, 73, 142, 206	0
1	N	306/309 (99%)	0.10	9 (2%)	55	50	44, 82, 146, 189	0
1	O	306/309 (99%)	0.20	12 (3%)	43	38	59, 96, 163, 215	0
1	P	306/309 (99%)	0.30	25 (8%)	14	13	61, 107, 170, 202	0
All	All	4896/4944 (99%)	0.07	141 (2%)	55	50	34, 82, 146, 294	0

The worst 5 of 141 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	245	THR	11.1
1	M	288	ALA	9.5
1	L	243	LYS	7.1
1	O	286	PHE	6.9
1	A	244	ALA	6.2

## 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 carbohydrates 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.