WORLDWIDE PDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Jan 30, 2017 – 10:33 AM EST

PDB ID	:	5FDK
Title	:	Crystal structure of RecU(D88N) in complex with palindromic DNA duplex
Authors	:	Khavnekar, S.; Rafferty, J.B.; Kale, A.
Deposited on	:	2015-12-16
Resolution	:	3.21 Å(reported)

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

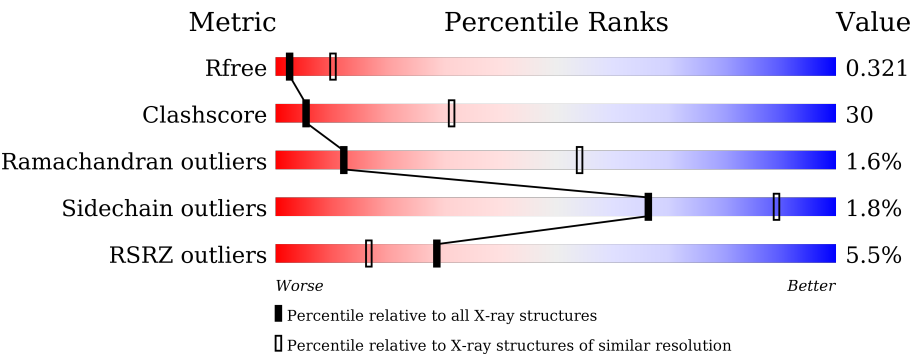|		
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20028442
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20028442

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 3.21 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		0.321
Clashscore		30
Ramachandran outliers		1.6%
Sidechain outliers		1.8%
RSRZ outliers		5.5%

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$_{free}$	91344	1124 (3.24-3.16)
Clashscore	102246	1024 (3.22-3.18)
Ramachandran outliers	100387	1004 (3.22-3.18)
Sidechain outliers	100360	1003 (3.22-3.18)
RSRZ outliers	91569	1129 (3.24-3.16)

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	199	5% / 54% 30% 5% 11%
1	B	199	% / 55% 26% ·· 17%
1	C	199	3% / 50% 35% ·· 13%
1	D	199	12% / 61% 35% ··
2	E	12	8% 58% 33%
2	F	12	17% 25% 58%

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	G	12	33% 33% 25% 8%
2	H	12	8% 25% 50% 17% 8%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 6138 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 Holliday junction resolvase RecU.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	177	Total	C	N	O	S	0	0	0
			1313	845	216	250	2			
1	B	166	Total	C	N	O	S	0	0	0
			1238	793	200	244	1			
1	C	174	Total	C	N	O	S	0	0	0
			1278	815	214	247	2			
1	D	199	Total	C	N	O	S	0	0	0
			1366	863	229	272	2			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	88	ASN	ASP	engineered mutation	UNP P39792
B	88	ASN	ASP	engineered mutation	UNP P39792
C	88	ASN	ASP	engineered mutation	UNP P39792
D	88	ASN	ASP	engineered mutation	UNP P39792

- Molecule 2 is a DNA chain called palindromic DNA.

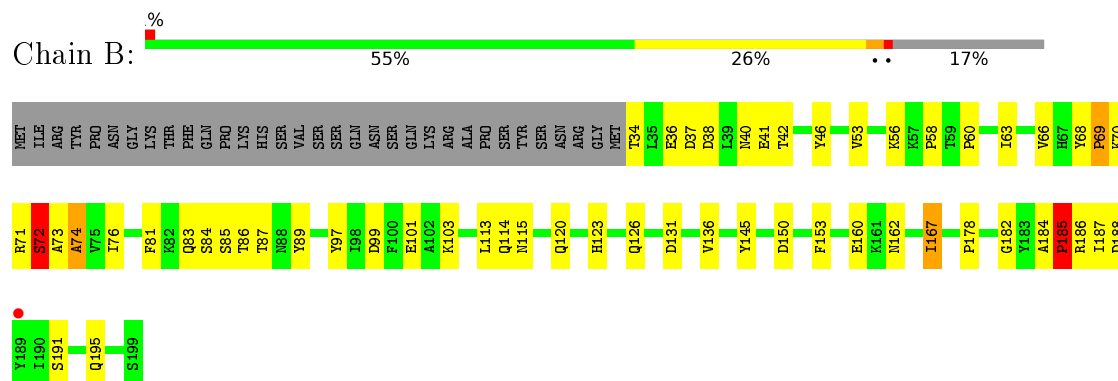| Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	12	Total	C	N	O	P	0	0	0
			246	117	45	72	12			
2	F	12	Total	C	N	O	P	0	0	0
			246	117	45	72	12			
2	G	11	Total	C	N	O	P	0	0	0
			225	107	40	67	11			
2	H	11	Total	C	N	O	P	0	0	0
			226	107	43	65	11			

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.
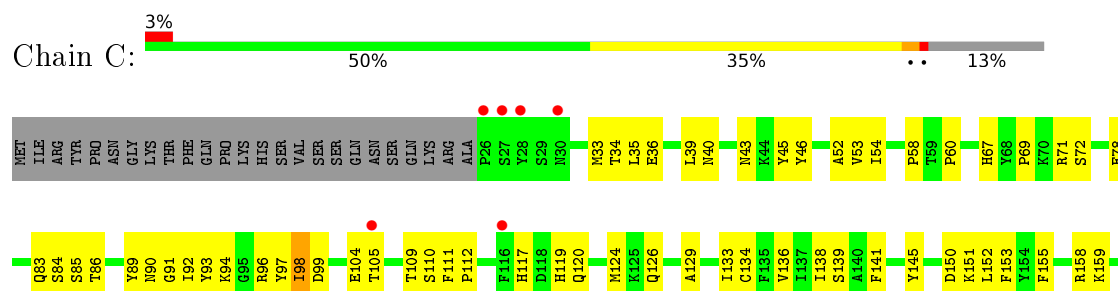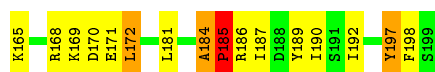
- Molecule 1: Holliday junction resolvase RecU

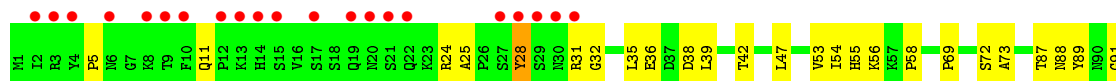
- Molecule 1: Holliday junction resolvase RecU

- Molecule 1: Holliday junction resolvase RecU

K165 R168 K169 D170 E171 L172 L181 A184 P185 R186 I187 D188 Y189 I190 S191 I192 Y197 F198 S199

• Molecule 1: Holliday junction resolvase RecU

Chain D:

12%

61%   35%

M1 I2 R3 Y4 P5 N6 G7 K8 T9 F10 Q11 P12 K13 H14 S15 V16 S17 S18 Q19 N20 S21 Q22 K23 R24 A25 P26 S27 Y28 S29 N30 R31 G32 L35 E36 D37 D38 L39 T42 I47 V53 I54 H55 K56 K57 P58 P69 S72 A73 T87 N88 Y89 N90 G91

I92 Y93 R96 Y97 I98 D99 F100 E101 E104 T109 S110 F111 P112 L113 H117 I121 M124 Q130 D131 G132 I133 C134 F135 V136 S139 A140 F141 D142 Y145 F146 L147 E148 K151 Y154 F155 R158 K159 G163 R164 K165 S166 I167 E171 E174 T175

A176 Y177 P178 I179 G182 Y183 A184 P185 R186 I187 D188 S191 I192 I193 E194 Y197 F198 S199

• Molecule 2: palindromic DNA

Chain E:

8%   58%   33%

A1 C2 G3 C4 A5 A6 T7 T8 G9 C10 G11 T12

• Molecule 2: palindromic DNA

Chain F:

17%   25%   58%

A1 C2 A5 A6 T7 T8 G9 C10 G11 T12

• Molecule 2: palindromic DNA

Chain G:

33%   33%   25%   8%

DA C2 G3 C4 T7 T8 G9 C10 G11 T12

• Molecule 2: palindromic DNA

Chain H:

8%

25%   50%   17%   8%

A1 C2 G3 C4 A5 A6 T7 T8 G9 C10 G11 DT

4   Data and refinement statistics (i)

Property	Value	Source
Space group	H 3 2	Depositor
Cell constants a, b, c, α, β, γ	144.45Å 144.45Å 310.34Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	48.69 – 3.21	Depositor
	48.69 – 3.21	EDS
% Data completeness	99.7 (48.69-3.21)	Depositor
(in resolution range)	99.8 (48.69-3.21)	EDS
R_{merge}	0.09	Depositor
R_{sym}	0.10	Depositor
$<I/\sigma(I)>$ [1]	1.75 (at 3.19Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.256 , 0.329	Depositor
	0.265 , 0.321	DCC
R_{free} test set	1062 reflections (5.13%)	DCC
Wilson B-factor (Å²)	119.6	Xtriage
Anisotropy	0.063	Xtriage
Bulk solvent k_{sol}(e/Å³), B_{sol}(Å²)	0.28 , 88.2	EDS
L-test for twinning[2]	$<|L|> = 0.48, <L^2> = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	6138	wwPDB-VP
Average B, all atoms (Å²)	82.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.42% 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 (i)

5.1 Standard geometry (i)

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	0.66	0/1347	0.96	7/1840 (0.4%)
1	B	0.67	1/1265 (0.1%)	1.11	6/1728 (0.3%)
1	C	0.57	0/1311	0.82	4/1794 (0.2%)
1	D	0.62	3/1398 (0.2%)	0.91	8/1923 (0.4%)
2	E	0.99	0/275	1.80	11/422 (2.6%)
2	F	0.87	0/275	1.95	18/422 (4.3%)
2	G	0.67	0/251	1.84	12/385 (3.1%)
2	H	0.69	0/253	1.84	13/388 (3.4%)
All	All	0.67	4/6375 (0.1%)	1.17	79/8902 (0.9%)

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	A	0	2
1	B	0	2
1	D	0	2
All	All	0	6

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	5	PRO	N-CD	-8.72	1.35	1.47
1	D	167	ILE	C-N	5.88	1.47	1.34
1	D	166	SER	C-N	5.74	1.47	1.34
1	B	185	PRO	N-CD	5.51	1.55	1.47

The worst 5 of 79 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	72	SER	CB-CA-C	18.39	145.04	110.10

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	72	SER	C-N-CA	-17.20	78.69	121.70
2	H	1	DA	O4'-C1'-N9	11.52	116.06	108.00
1	B	72	SER	N-CA-C	-11.30	80.48	111.00
1	A	186	ARG	N-CA-C	-11.11	81.00	111.00

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	28	TYR	Peptide
1	A	33	MET	Peptide
1	B	184	ALA	Peptide
1	B	72	SER	Peptide
1	D	11	GLN	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	1313	0	1119	89	0
1	B	1238	0	1081	59	0
1	C	1278	0	1046	64	0
1	D	1366	0	1030	98	1
2	E	246	0	136	9	0
2	F	246	0	136	14	0
2	G	225	0	125	3	1
2	H	226	0	124	5	0
All	All	6138	0	4797	323	1

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 30.

The worst 5 of 323 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:69:PRO:CG	1:B:73:ALA:HA	1.43	1.44

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:69:PRO:CD	1:B:74:ALA:H	1.30	1.41
1:B:69:PRO:HG2	1:B:73:ALA:CA	1.62	1.29
1:B:69:PRO:HG2	1:B:72:SER:O	1.38	1.23
1:D:69:PRO:HB2	1:D:73:ALA:CB	1.76	1.16

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:109:THR:CG2	2:G:7:DT:OP2[7_445]	2.03	0.17

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	175/199 (88%)	147 (84%)	23 (13%)	5 (3%)	6	36
1	B	164/199 (82%)	146 (89%)	16 (10%)	2 (1%)	16	60
1	C	172/199 (86%)	139 (81%)	32 (19%)	1 (1%)	30	75
1	D	197/199 (99%)	160 (81%)	34 (17%)	3 (2%)	13	55
All	All	708/796 (89%)	592 (84%)	105 (15%)	11 (2%)	12	54

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	184	ALA
1	B	185	PRO
1	C	185	PRO
1	A	48	THR
1	B	69	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	115/181 (64%)	114 (99%)	1 (1%)	84	95
1	B	114/181 (63%)	113 (99%)	1 (1%)	84	95
1	C	108/181 (60%)	104 (96%)	4 (4%)	41	79
1	D	99/181 (55%)	97 (98%)	2 (2%)	63	88
All	All	436/724 (60%)	428 (98%)	8 (2%)	66	89

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	98	ILE
1	D	154	TYR
1	C	198	PHE
1	C	96	ARG
1	C	197	TYR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	43	ASN
1	A	90	ASN
1	A	123	HIS
1	B	114	GLN
1	D	55	HIS

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 (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 (i)

6.1 Protein, DNA and RNA chains (i)

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(Å^2)	Q<0.9
1	A	177/199 (88%)	0.14	10 (5%)	28	16	51, 68, 100, 146	9 (5%)
1	B	166/199 (83%)	0.11	1 (0%)	90	84	52, 71, 104, 117	0
1	C	174/199 (87%)	0.12	6 (3%)	49	34	50, 81, 108, 126	5 (2%)
1	D	199/199 (100%)	0.41	24 (12%)	6	3	49, 90, 115, 143	31 (15%)
2	E	12/12 (100%)	0.29	0	100	100	59, 79, 112, 113	0
2	F	12/12 (100%)	-0.06	0	100	100	62, 78, 104, 111	0
2	G	11/12 (91%)	-0.05	0	100	100	111, 123, 152, 153	0
2	H	11/12 (91%)	0.97	1 (9%)	11	6	105, 123, 162, 163	0
All	All	762/844 (90%)	0.21	42 (5%)	29	16	49, 78, 116, 163	45 (5%)

The worst 5 of 42 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	21	SER	6.4
1	D	9	THR	5.7
1	A	31	ARG	4.9
1	D	22	GLN	4.3
1	D	27	SER	4.2

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.