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 (i)

May 25, 2020 – 11:05 pm BST

PDB ID	:	1DVK
Title	:	CRYSTAL STRUCTURE OF THE FUNCTIONAL DOMAIN OF THE SPLICING FACTOR PRP18
Authors	:	Jiang, J.; Horowitz, D.S.; Xu, R.M.
Deposited on	:	2000-01-21
Resolution	:	2.15 Å(reported)

The following versions of software and data (see references (1)) were used in the production of this report:

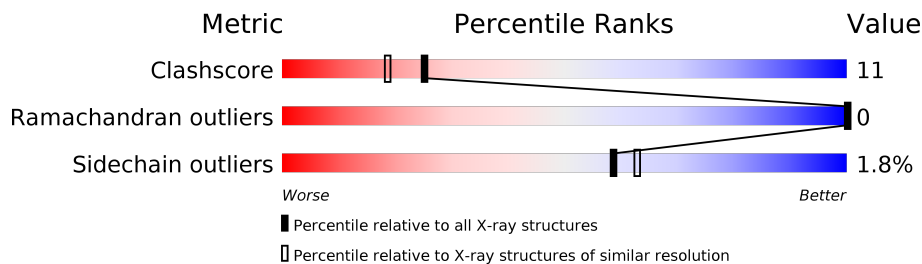|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 2.15 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
Clashscore		11
Ramachandran outliers		0
Sidechain outliers		1.8%

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))
Clashscore	141614	1585 (2.16-2.16)
Ramachandran outliers	138981	1560 (2.16-2.16)
Sidechain outliers	138945	1559 (2.16-2.16)

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 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%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	173	62% 22% • 14%
1	B	173	67% 20% 13%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2646 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 PRP18.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	149	Total	C	N	O	S	0	0	0
			1210	785	205	215	5			
1	B	150	Total	C	N	O	S	0	0	0
			1217	789	206	217	5			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	79	MET	GLU	SEE REMARK 999	UNP P33411
B	79	MET	GLU	SEE REMARK 999	UNP P33411

- Molecule 2 is water.

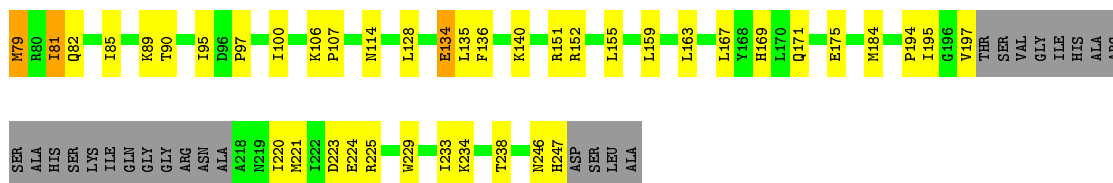| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	107	Total	O	0	0
			107	107		
2	B	112	Total	O	0	0
			112	112		

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 geometry. 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. 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.

Note EDS was not executed.

• Molecule 1: PRP18

Chain A:

62%   22%   •   14%

M79 R80 I81 Q82 I85 K89 T90 I95 D96 P97 I100 K106 P107 N114 L128 E134 L135 F136 K140 R151 R152 L155 L159 L163 L167 Y168 H169 L170 Q171 E175 M184 P194 I195 G196 V197 THR SER VAL GLY ILE HIS ALA ARG

SER ALA HIS SER LYS ILE GLN GLY GLY ARG ASN ALA A218 N219 I220 M221 I222 D223 E224 R225 W229 I233 K234 T238 N246 H247 ASP SER LEU ALA

• Molecule 1: PRP18

Chain B:

67%   20%   13%

M79 R80 I81 Q82 Q87 I95 K106 P107 W124 Y131 H132 L137 K140 K141 A142 L143 F144 P145 R151 P157 L160 I161 H169 L170 Q171 Q172 P173 K174 E175 I176 N177 L178 Q181 I195 T198 SER VAL GLY ILE HIS ALA ARG SER ALA HIS SER

LYS ILE GLN GLY GLY ARG ASN ALA A218 E224 L228 W229 I233 F239 H247 ASP SER LEU ALA

4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	49.94Å 79.17Å 171.19Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.00 – 2.15	Depositor
% Data completeness (in resolution range)	92.6 (40.00-2.15)	Depositor
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	CNS	Depositor
R, R_{free}	0.203 , 0.252	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	2646	wwPDB-VP
Average B, all atoms (Å^2)	32.0	wwPDB-VP

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	0.29	0/1235	0.50	0/1675
1	B	0.29	0/1242	0.49	0/1685
All	All	0.29	0/2477	0.50	0/3360

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 (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	1210	0	1257	32	0
1	B	1217	0	1264	21	0
2	A	107	0	0	2	0
2	B	112	0	0	2	0
All	All	2646	0	2521	53	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 11.

The worst 5 of 53 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:95:ILE:H	1:B:171:GLN:HE22	1.33	0.77

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:177:ASN:O	1:B:181:GLN:HG3	1.91	0.69
1:B:169:HIS:CE1	1:B:195:ILE:H	2.13	0.67
1:A:169:HIS:CE1	1:A:195:ILE:H	2.13	0.66
1:A:169:HIS:HE1	1:A:195:ILE:H	1.47	0.63

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	145/173 (84%)	141 (97%)	4 (3%)	0	100	100
1	B	146/173 (84%)	141 (97%)	5 (3%)	0	100	100
All	All	291/346 (84%)	282 (97%)	9 (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	135/152 (89%)	131 (97%)	4 (3%)	41	40
1	B	136/152 (90%)	135 (99%)	1 (1%)	84	89
All	All	271/304 (89%)	266 (98%)	5 (2%)	59	63

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	79	MET
1	A	81	ILE
1	A	134	GLU
1	A	184	MET
1	B	143	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	118	HIS
1	B	153	ASN
1	B	181	GLN
1	B	114	ASN
1	B	177	ASN

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)

EDS was not executed - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains (i)

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates (i)

EDS was not executed - this section is therefore empty.

6.4 Ligands (i)

EDS was not executed - this section is therefore empty.

6.5 Other polymers (i)

EDS was not executed - this section is therefore empty.