WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Nov 7, 2023 – 12:09 PM JST

PDB ID	:	8J0H
Title	:	Crystal structure of the fission yeast Rex1BD protein(C4H3.06)
Authors	:	Li, J.; Sun, W.; Chen, Y.
Deposited on	:	2023-04-11
Resolution	:	3.38 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) 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 (i)) were used in the production of this report:

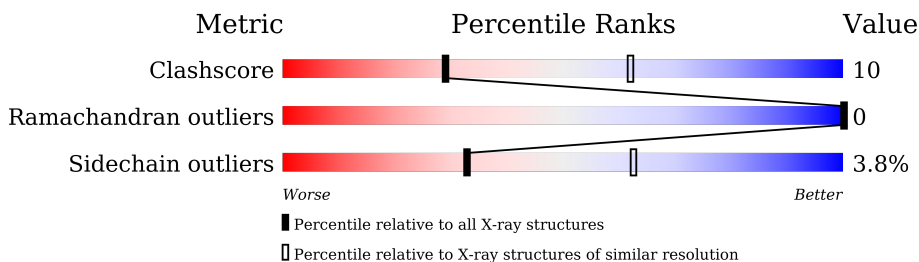|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 3.38 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		10
Ramachandran outliers		0
Sidechain outliers		3.8%

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	1762 (3.46-3.30)
Ramachandran outliers	138981	1732 (3.46-3.30)
Sidechain outliers	138945	1731 (3.46-3.30)

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 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%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	142	69% 25% 6%
1	B	142	70% 23% 7%
1	C	142	71% 20% • 8%
1	D	142	63% 26% • 9%
1	E	142	70% 21% • 8%
1	F	142	75% 14% • 8%
1	G	142	71% 21% 8%
1	H	142	65% 22% • 9%

2 Entry composition ⓘ

There is only 1 type of molecule in this entry. The entry contains 8308 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 Uncharacterized protein C4H3.06.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	134	Total	C	N	O	Se	0	0	0
			1063	657	181	222	3			
1	B	132	Total	C	N	O	Se	0	0	0
			1048	646	179	220	3			
1	C	131	Total	C	N	O	Se	0	0	0
			1040	640	178	219	3			
1	D	129	Total	C	N	O	Se	0	0	0
			1023	631	176	213	3			
1	E	130	Total	C	N	O	Se	0	0	0
			1032	636	177	216	3			
1	F	131	Total	C	N	O	Se	0	0	0
			1040	640	178	219	3			
1	G	131	Total	C	N	O	Se	0	0	0
			1040	640	178	219	3			
1	H	129	Total	C	N	O	Se	0	0	0
			1022	628	174	217	3			

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-4	GLY	-	expression tag	UNP Q10214
A	-3	PRO	-	expression tag	UNP Q10214
A	-2	LEU	-	linker	UNP Q10214
A	-1	GLY	-	linker	UNP Q10214
A	0	SER	-	linker	UNP Q10214
B	-4	GLY	-	expression tag	UNP Q10214
B	-3	PRO	-	expression tag	UNP Q10214
B	-2	LEU	-	linker	UNP Q10214
B	-1	GLY	-	linker	UNP Q10214
B	0	SER	-	linker	UNP Q10214
C	-4	GLY	-	expression tag	UNP Q10214
C	-3	PRO	-	expression tag	UNP Q10214
C	-2	LEU	-	linker	UNP Q10214

Continued on next page...

Continued from previous page...

| Chain	Residue	Modelled	Actual	Comment	Reference
C	-1	GLY	-	linker	UNP Q10214
C	0	SER	-	linker	UNP Q10214
D	-4	GLY	-	expression tag	UNP Q10214
D	-3	PRO	-	expression tag	UNP Q10214
D	-2	LEU	-	linker	UNP Q10214
D	-1	GLY	-	linker	UNP Q10214
D	0	SER	-	linker	UNP Q10214
E	-4	GLY	-	expression tag	UNP Q10214
E	-3	PRO	-	expression tag	UNP Q10214
E	-2	LEU	-	linker	UNP Q10214
E	-1	GLY	-	linker	UNP Q10214
E	0	SER	-	linker	UNP Q10214
F	-4	GLY	-	expression tag	UNP Q10214
F	-3	PRO	-	expression tag	UNP Q10214
F	-2	LEU	-	linker	UNP Q10214
F	-1	GLY	-	linker	UNP Q10214
F	0	SER	-	linker	UNP Q10214
G	-4	GLY	-	expression tag	UNP Q10214
G	-3	PRO	-	expression tag	UNP Q10214
G	-2	LEU	-	linker	UNP Q10214
G	-1	GLY	-	linker	UNP Q10214
G	0	SER	-	linker	UNP Q10214
H	-4	GLY	-	expression tag	UNP Q10214
H	-3	PRO	-	expression tag	UNP Q10214
H	-2	LEU	-	linker	UNP Q10214
H	-1	GLY	-	linker	UNP Q10214
H	0	SER	-	linker	UNP Q10214

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. 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.
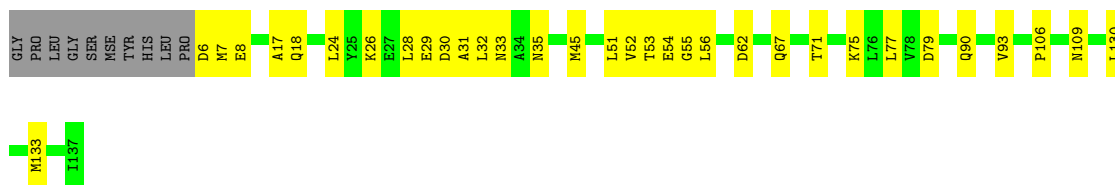
Note EDS was not executed.

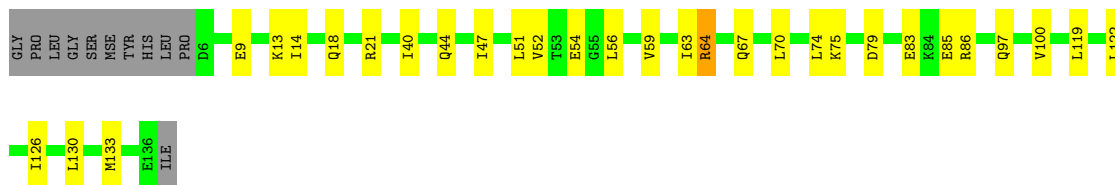• Molecule 1: Uncharacterized protein C4H3.06

Chain A:
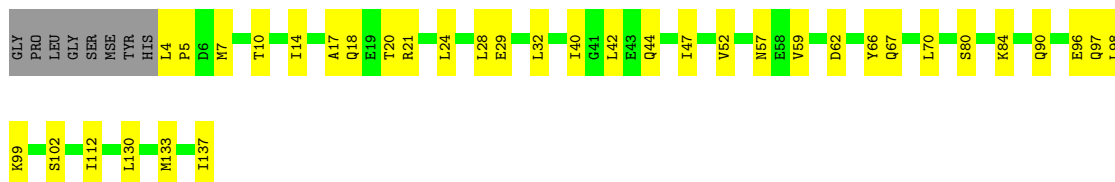
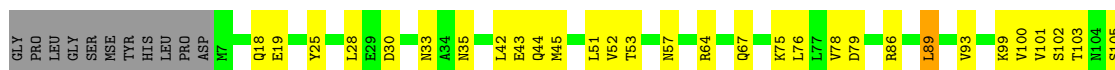• Molecule 1: Uncharacterized protein C4H3.06

Chain B:

• Molecule 1: Uncharacterized protein C4H3.06

Chain C:

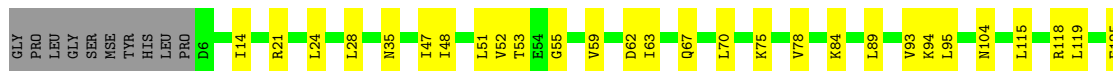
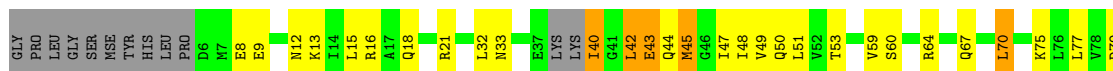• Molecule 1: Uncharacterized protein C4H3.06

Chain D:

P106 V107 E108 N109 T110 L122 E125 L130 Q131 N132 M133 K134 D135 GLU ILE

• Molecule 1: Uncharacterized protein C4H3.06

Chain E: 70% 21% • 8%

GLY PRO LEU GLY SER MSE TYR HIS LEU PRO ASP M7 K13 I14 A17 R21 L28 I40 G41 L42 V52 T53 E54 V59 I63 L70 T71 K72 E73 L74 K75 L76 L77 V78 D79 S80 L81 K84 K88 T92 V93 T103 E108 R118 L119

L122 S123 M133 E136 ILE

• Molecule 1: Uncharacterized protein C4H3.06

Chain F: 75% 14% • 8%

GLY PRO LEU GLY SER MSE TYR HIS LEU PRO D6 M7 Q18 R21 Y25 K26 L32 M45 Q50 V59 T71 E85 L89 K94 L95 E96 Q97 L98 V101 S105 P106 T110 L115 L130 Q131 D135 E136 ILE

• Molecule 1: Uncharacterized protein C4H3.06

Chain G: 71% 21% 8%

GLY PRO LEU GLY SER MSE TYR HIS LEU PRO D6 I14 R21 L24 L28 N35 I47 I48 L51 V52 T53 E54 G55 V59 D62 I63 Q67 L70 K75 V78 K84 L89 V93 K94 L95 N104 L115 R118 L119 E125

I129 M133 E136 ILE

• Molecule 1: Uncharacterized protein C4H3.06

Chain H: 65% 22% • 9%

GLY PRO LEU GLY SER MSE TYR HIS LEU PRO D6 M7 E8 E9 N12 K13 I14 L15 R16 A17 Q18 R21 L32 N33 E37 LYS LYS I40 G41 L42 E43 Q44 M45 G46 I47 I48 V49 Q50 L51 V52 T53 V59 S60 R64 Q67 L70 K75 L76 L77 V78 D79

E85 Q90 L95 E116 I129 L130 M133 E136 ILE

4 Data and refinement statistics ⓘ

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

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	159.01Å 138.48Å 50.90Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.47 – 3.38	Depositor
% Data completeness (in resolution range)	99.2 (48.47-3.38)	Depositor
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	PHENIX (1.14_3260: ???)	Depositor
R, R_{free}	0.242 , 0.296	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	8308	wwPDB-VP
Average B, all atoms (Å^2)	92.0	wwPDB-VP

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	0.22	0/1063	0.38	0/1424
1	B	0.22	0/1047	0.36	0/1401
1	C	0.23	0/1039	0.40	0/1390
1	D	0.27	0/1023	0.39	0/1370
1	E	0.23	0/1032	0.38	0/1382
1	F	0.27	0/1039	0.38	0/1390
1	G	0.22	0/1039	0.37	0/1390
1	H	0.27	0/1020	0.43	0/1365
All	All	0.24	0/8302	0.39	0/11112

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	A	1063	0	1101	25	0
1	B	1048	0	1083	24	0
1	C	1040	0	1072	23	0
1	D	1023	0	1062	21	0
1	E	1032	0	1068	24	0
1	F	1040	0	1072	18	0
1	G	1040	0	1072	20	0
1	H	1022	0	1045	27	0

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	8308	0	8575	162	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 10.

The worst 5 of 162 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:12:ASN:HB3	1:H:16:ARG:NH1	1.75	0.99
1:G:47:ILE:O	1:G:51:LEU:HD22	1.73	0.86
1:H:12:ASN:HB3	1:H:16:ARG:HH11	1.38	0.86
1:E:78:VAL:HG22	1:E:133:MSE:CE	2.07	0.85
1:H:32:LEU:HD23	1:H:45:MSE:SE	2.28	0.84

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	132/142 (93%)	130 (98%)	2 (2%)	0	100	100
1	B	130/142 (92%)	130 (100%)	0	0	100	100
1	C	129/142 (91%)	128 (99%)	1 (1%)	0	100	100
1	D	127/142 (89%)	126 (99%)	1 (1%)	0	100	100
1	E	128/142 (90%)	128 (100%)	0	0	100	100
1	F	129/142 (91%)	127 (98%)	2 (2%)	0	100	100
1	G	129/142 (91%)	128 (99%)	1 (1%)	0	100	100
1	H	125/142 (88%)	123 (98%)	2 (2%)	0	100	100
All	All	1029/1136 (91%)	1020 (99%)	9 (1%)	0	100	100

There are no Ramachandran outliers to report.

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	124/126 (98%)	123 (99%)	1 (1%)	81	91
1	B	122/126 (97%)	120 (98%)	2 (2%)	62	81
1	C	121/126 (96%)	119 (98%)	2 (2%)	60	80
1	D	119/126 (94%)	109 (92%)	10 (8%)	11	37
1	E	120/126 (95%)	117 (98%)	3 (2%)	47	73
1	F	121/126 (96%)	113 (93%)	8 (7%)	16	47
1	G	121/126 (96%)	118 (98%)	3 (2%)	47	73
1	H	119/126 (94%)	111 (93%)	8 (7%)	16	47
All	All	967/1008 (96%)	930 (96%)	37 (4%)	33	62

5 of 37 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	104	ASN
1	H	70	LEU
1	H	8	GLU
1	H	42	LEU
1	D	107	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 ⓘ

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates ⓘ

EDS was not executed - this section is therefore empty.

6.4 Ligands ⓘ

EDS was not executed - this section is therefore empty.

6.5 Other polymers ⓘ

EDS was not executed - this section is therefore empty.