wwPDB X-ray Structure Validation Summary Report (i)

Oct 9, 2023 – 10:37 PM EDT

PDB ID	:	7MI1
Title	:	X-ray structure of yeast dynein motor domain in the presence of a pyrazolo-p yrimidinone-based compound (compound 20)
Authors	:	Santarossa, C.C.; Ekiert, D.C.; Bhabha, G.; Kapoor, T.M.
Deposited on	:	2021-04-16
Resolution	:	4.50 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 (1)) were used in the production of this report:

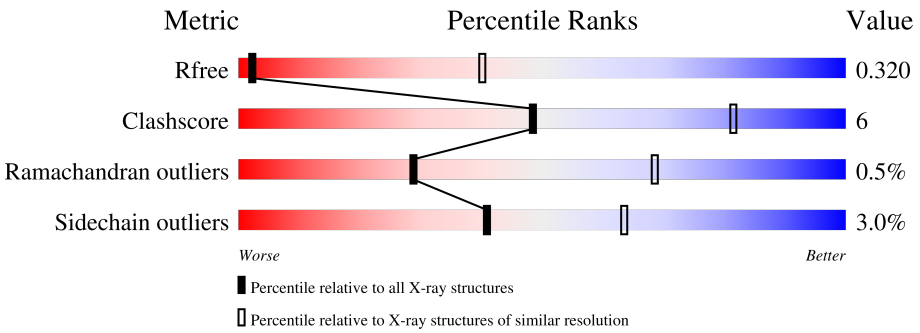
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 4.50 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}	130704	1055 (5.20-3.80)
Clashscore	141614	1123 (5.20-3.80)
Ramachandran outliers	138981	1069 (5.20-3.80)
Sidechain outliers	138945	1050 (5.20-3.80)

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%

Mol	Chain	Length	Quality of chain
1	A	2661	81% 16% ••

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 42528 atoms, of which 21316 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 Chimera protein of Dynein and Endolysin.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	2628	42528	13593	21316	3538	3984	97	0	0	0

There are 31 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1363	GLY	-	expression tag	UNP P36022
A	1849	GLN	GLU	engineered mutation	UNP P36022
A	3120	GLY	-	linker	UNP P36022
A	3121	SER	-	linker	UNP P36022
A	3122	GLY	-	linker	UNP P36022
A	3123	SER	-	linker	UNP P36022
A	3124	GLY	-	linker	UNP P36022
A	3125	SER	-	linker	UNP P36022
A	3178	THR	CYS	conflict	UNP D9IEF7
A	3221	ALA	CYS	conflict	UNP D9IEF7
A	3286	GLY	-	linker	UNP D9IEF7
A	3287	SER	-	linker	UNP D9IEF7
A	3288	GLY	-	linker	UNP D9IEF7
A	3289	SER	-	linker	UNP D9IEF7
A	3290	GLY	-	linker	UNP D9IEF7
A	3291	SER	-	linker	UNP D9IEF7
A	3742	ASP	ASN	conflict	UNP P36022
A	3895	VAL	PHE	conflict	UNP P36022
A	4072	ASP	ASN	conflict	UNP P36022
A	4093	GLY	-	expression tag	UNP P36022
A	4094	SER	-	expression tag	UNP P36022
A	4095	GLY	-	expression tag	UNP P36022
A	4096	SER	-	expression tag	UNP P36022
A	4097	GLY	-	expression tag	UNP P36022
A	4098	SER	-	expression tag	UNP P36022
A	4099	HIS	-	expression tag	UNP P36022
A	4100	HIS	-	expression tag	UNP P36022

Continued on next page...

Continued from previous page...

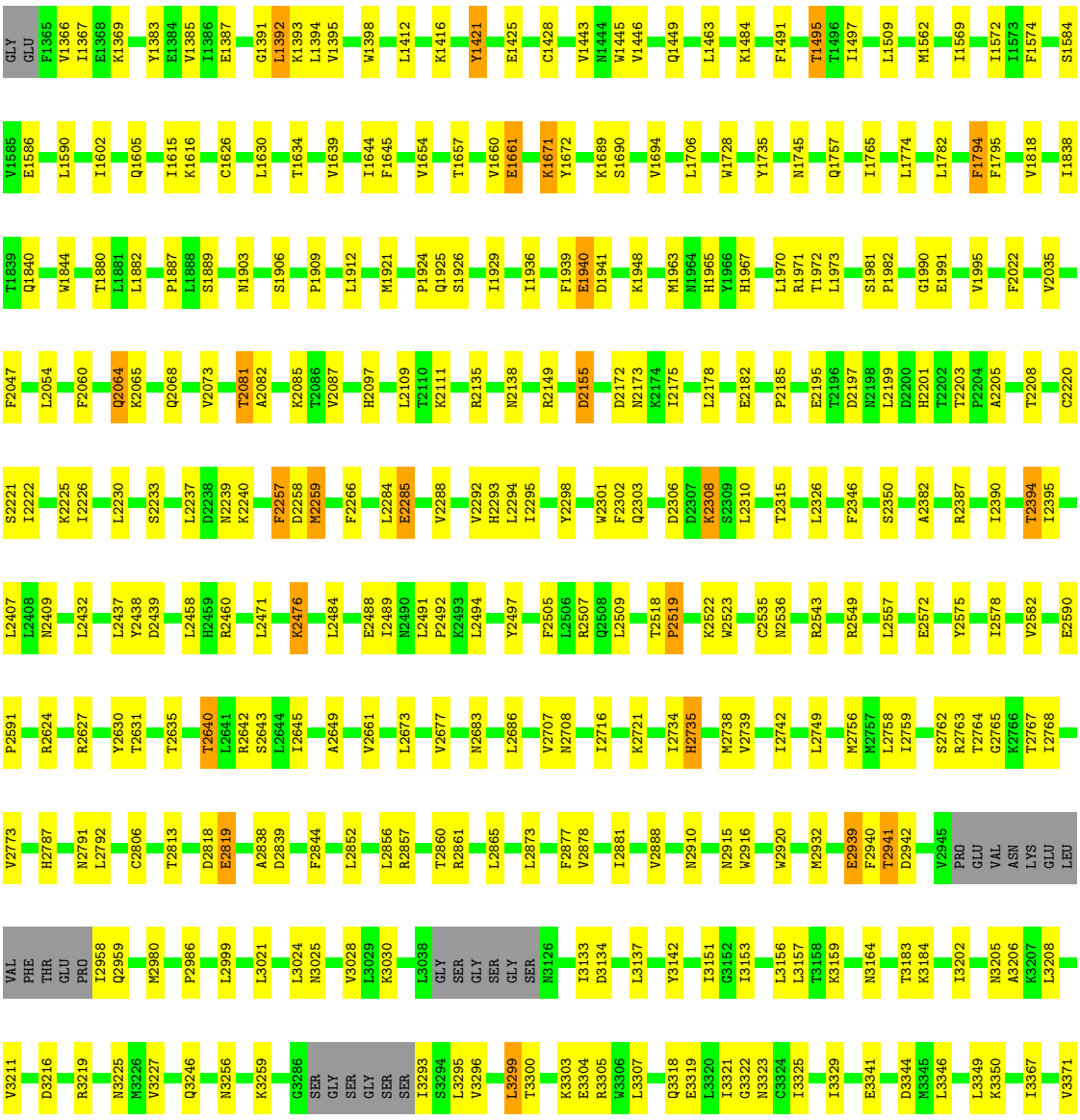
Chain	Residue	Modelled	Actual	Comment	Reference
A	4101	HIS	-	expression tag	UNP P36022
A	4102	HIS	-	expression tag	UNP P36022
A	4103	HIS	-	expression tag	UNP P36022
A	4104	HIS	-	expression tag	UNP P36022

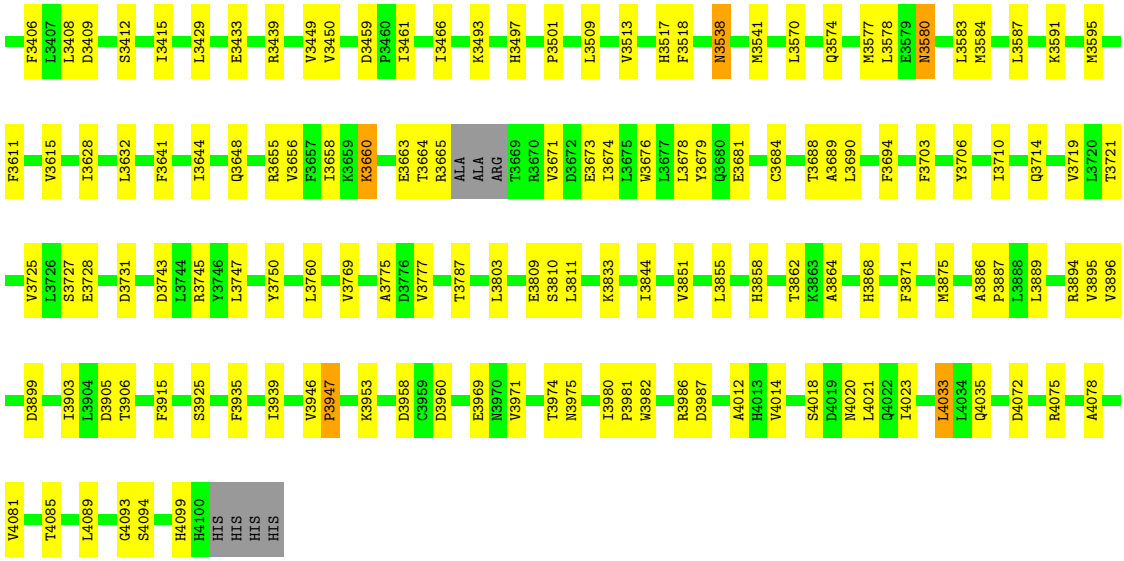
3 Residue-property plots (i)

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.

- Molecule 1: Chimera protein of Dynein and Endolysin

Chain A: 81% 16% ••




F3406 L3407 L3408 D3409 S3412 I3415 L3429 E3433 R3439 V3449 V3450 D3459 P3460 I3461 I3466 K3493 H3497 P3501 L3509 V3513 H3517 F3518 N3538 M3541 L3570 Q3574 M3577 L3578 E3579 N3580 L3583 M3584 L3587 K3591 M3595
F3611 V3615 I3628 L3632 F3641 I3644 Q3648 R3655 V3656 F3657 I3658 K3659 K3660 E3663 T3664 R3665 ALA ALA ARG T3669 R3670 V3671 D3672 E3673 I3674 L3675 W3676 L3677 L3678 Y3679 Q3680 E3681 C3684 T3688 A3689 L3690 F3694 F3703 Y3706 I3710 Q3714 V3719 L3720 T3721
V3725 L3726 S3727 E3728 D3731 D3743 L3744 R3745 Y3746 L3747 Y3750 L3760 V3769 A3775 D3776 V3777 T3787 L3803 E3809 S3810 L3811 K3833 I3844 V3851 L3855 H3858 T3862 K3863 A3864 H3868 F3871 M3875 A3886 P3887 L3888 L3889 R3894 V3895 V3896
D3899 I3903 L3904 D3905 T3906 F3915 S3925 F3935 I3939 V3946 P3947 K3953 D3958 C3959 D3960 E3969 N3970 V3971 T3974 N3975 I3980 P3981 W3982 R3986 D3987 A4012 H4013 V4014 S4018 D4019 N4020 L4021 Q4022 I4023 L4033 L4034 Q4035 D4072 R4075 A4078
V4081 T4085 L4089 G4093 S4094 H4099 H4100 HIS HIS HIS HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 2 21 21	Depositor
Cell constants a, b, c, α, β, γ	135.06Å 157.92Å 179.31Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.66 – 4.50 47.66 – 4.50	Depositor EDS
% Data completeness (in resolution range)	99.5 (47.66-4.50) 99.5 (47.66-4.50)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.14 (at 4.45Å)	Xtriage
Refinement program	PHENIX 1.17.1_3660	Depositor
R, R_{free}	0.276 , 0.320 0.276 , 0.320	Depositor DCC
R_{free} test set	1997 reflections (8.61%)	wwPDB-VP
Wilson B-factor (Å^2)	183.0	Xtriage
Anisotropy	0.178	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 122.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.39$, $< L^2 > = 0.22$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.87	EDS
Total number of atoms	42528	wwPDB-VP
Average B, all atoms (Å^2)	228.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.50% 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	0.26	0/21632	0.44	1/29225 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	2941	THR	N-CA-C	5.17	124.97	111.00

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	21212	21316	21315	257	0
All	All	21212	21316	21315	257	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 6.

The worst 5 of 257 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:3665:ARG:NH1	1:A:3673:GLU:OE2	1.97	0.97
1:A:2939:GLU:OE1	1:A:3318:GLN:NE2	2.05	0.89

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:2624:ARG:NH2	1:A:2910:ASN:O	2.08	0.86
1:A:2064:GLN:OE1	1:A:2065:LYS:N	2.15	0.79
1:A:2677:VAL:HG11	1:A:2686:LEU:HD21	1.64	0.77

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	2618/2661 (98%)	2464 (94%)	141 (5%)	13 (0%)	29 68

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	1940	GLU
1	A	3810	SER
1	A	4099	HIS
1	A	1391	GLY
1	A	2259	MET

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	2381/2406 (99%)	2309 (97%)	72 (3%)	41 63

5 of 72 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	3745	ARG
1	A	4033	LEU
1	A	3750	TYR
1	A	3939	ILE
1	A	2295	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	1899	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 monosaccharides 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)

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers (i)

Unable to reproduce the depositors R factor - this section is therefore empty.