W O R L D W I D E

wwPDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

May 14, 2020 – 08:59 am BST

PDB ID	:	4J7B
Title	:	Crystal structure of polo-like kinase 1
Authors	:	Xu, J.; Shen, C.; Quan, J.; Wang, T.
Deposited on	:	2013-02-13
Resolution	:	2.30 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)	:	1.13
EDS	:	2.11
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)	:	2.11

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 2.30 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.
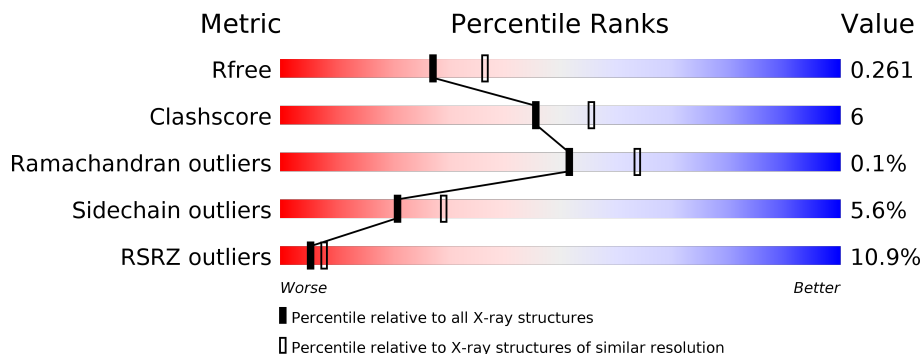
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	5042 (2.30-2.30)
Clashscore	141614	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)
RSRZ outliers	127900	4938 (2.30-2.30)

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 >=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% 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	297	9% / 81% 13% · ·
1	D	297	16% / 78% 15% · 5%
2	B	237	5% / 76% 13% · 10%
2	E	237	5% / 75% 14% · 10%
3	C	58	16% / 55% 28% · 16%
3	F	58	19% / 64% 19% · 16%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 9170 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 Polo-like kinase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	284	Total 2287	C 1470	N 401	O 400	S 16	0	0	0
1	D	281	Total 2265	C 1456	N 397	O 397	S 15	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	16	MET	-	EXPRESSION TAG	UNP Q6DRK7
A	17	GLY	-	EXPRESSION TAG	UNP Q6DRK7
A	162	ASN	ASP	ENGINEERED MUTATION	UNP Q6DRK7
A	192	PHE	GLU	ENGINEERED MUTATION	UNP Q6DRK7
D	16	MET	-	EXPRESSION TAG	UNP Q6DRK7
D	17	GLY	-	EXPRESSION TAG	UNP Q6DRK7
D	162	ASN	ASP	ENGINEERED MUTATION	UNP Q6DRK7
D	192	PHE	GLU	ENGINEERED MUTATION	UNP Q6DRK7

- Molecule 2 is a protein called Polo-like kinase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	213	Total 1737	C 1109	N 295	O 320	S 13	0	0	0
2	E	213	Total 1737	C 1109	N 295	O 320	S 13	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	359	MET	-	EXPRESSION TAG	UNP Q6DRK7
B	384	GLY	ASP	ENGINEERED MUTATION	UNP Q6DRK7
E	359	MET	-	EXPRESSION TAG	UNP Q6DRK7
E	384	GLY	ASP	ENGINEERED MUTATION	UNP Q6DRK7

- Molecule 3 is a protein called 205 kDa microtubule-associated protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	49	392	240	66	84	2	0	0	0
			Total	C	N	O	S			
3	F	49	392	240	66	84	2	0	0	0

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	268	MET	-	EXPRESSION TAG	UNP P23226
C	269	GLY	-	EXPRESSION TAG	UNP P23226
C	270	HIS	-	EXPRESSION TAG	UNP P23226
C	271	HIS	-	EXPRESSION TAG	UNP P23226
C	272	HIS	-	EXPRESSION TAG	UNP P23226
C	273	HIS	-	EXPRESSION TAG	UNP P23226
C	274	HIS	-	EXPRESSION TAG	UNP P23226
C	275	HIS	-	EXPRESSION TAG	UNP P23226
F	268	MET	-	EXPRESSION TAG	UNP P23226
F	269	GLY	-	EXPRESSION TAG	UNP P23226
F	270	HIS	-	EXPRESSION TAG	UNP P23226
F	271	HIS	-	EXPRESSION TAG	UNP P23226
F	272	HIS	-	EXPRESSION TAG	UNP P23226
F	273	HIS	-	EXPRESSION TAG	UNP P23226
F	274	HIS	-	EXPRESSION TAG	UNP P23226
F	275	HIS	-	EXPRESSION TAG	UNP P23226

- Molecule 4 is water.

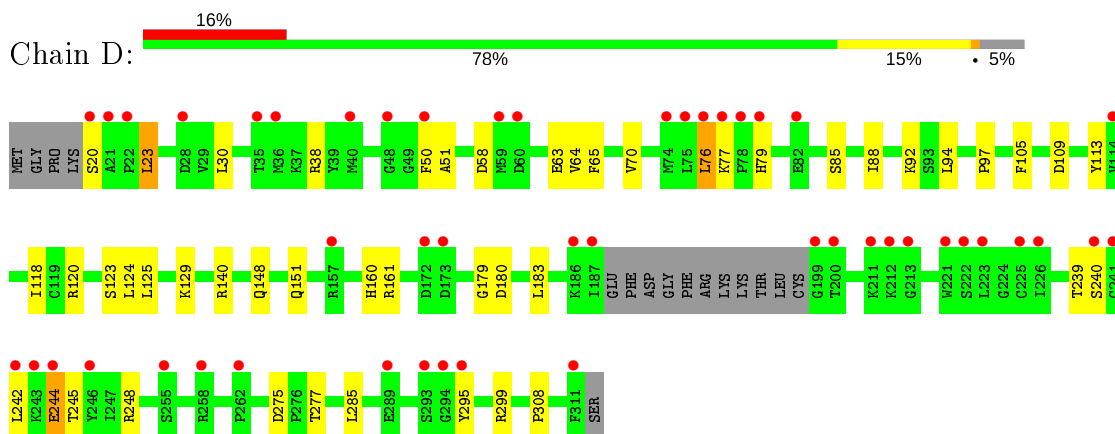| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	94	Total 94	O 94	0	0
4	B	128	Total 128	O 128	0	0
4	C	17	Total 17	O 17	0	0
4	D	54	Total 54	O 54	0	0
4	E	56	Total 56	O 56	0	0
4	F	11	Total 11	O 11	0	0

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 the various outlier classes 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.
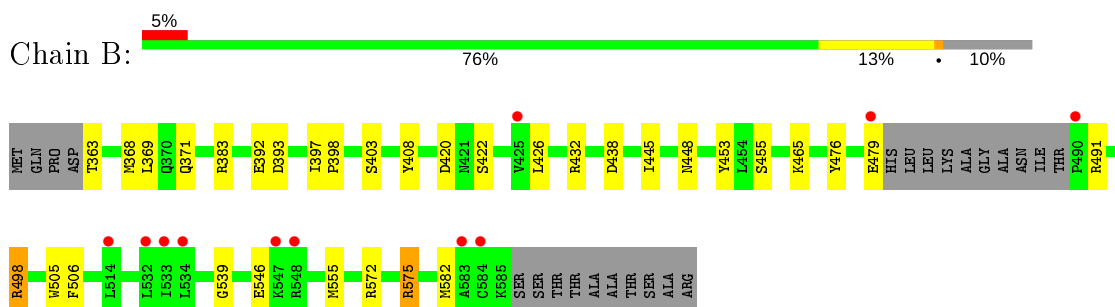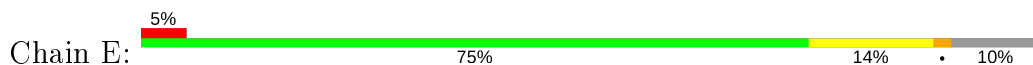
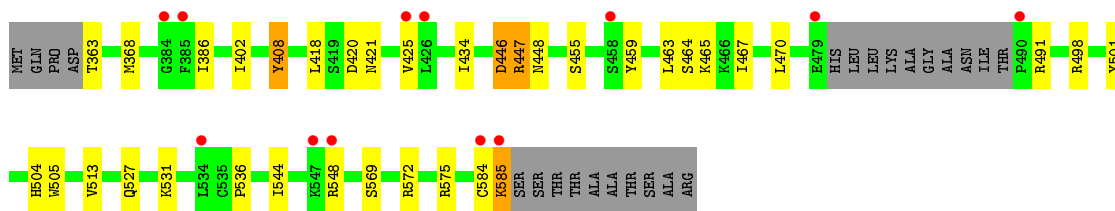- Molecule 1: Polo-like kinase

Chain A:

- Molecule 1: Polo-like kinase

Chain D:

- Molecule 2: Polo-like kinase

Chain B:

- Molecule 2: Polo-like kinase

Chain E:

MET GLN PRO ASP T363 M368 G384 F385 I386 I402 Y408 L418 S419 D420 N421 V425 L426 I434 D446 R447 N448 S455 S458 Y459 L463 S464 K465 K466 I467 L470 E479 HIS LEU LEU LYS ALA GLY ALA ASN ILE THR P490 R491 R498 Y501

H504 W505 V513 Q527 K531 L534 C535 P536 I544 K547 R548 S569 R572 R575 C584 K585 SER SER THR THR ALA ALA THR SER ALA ARG

• Molecule 3: 205 kDa microtubule-associated protein

Chain C:

16%

55%   28%   •   16%

MET GLY HIS HIS HIS HIS HIS HIS L276 D277 D278 L279 S283 P284 R285 K286 E287 F288 A289 R290 D294 R302 E303 I306 D309 M310 R311 P312 H313 E314 L315 E316 Q317 E318 S319 D320 T321 A324 GLY

• Molecule 3: 205 kDa microtubule-associated protein

Chain F:

19%

64%   19%   •   16%

MET GLY HIS HIS HIS HIS HIS HIS L276 R285 K286 E287 F288 A289 R290 I291 I296 E307 L315 E316 Q317 E318 S319 D320 T321 F322 G323 A324 GLY

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	47.68Å 57.45Å 125.74Å 89.15° 89.84° 72.38°	Depositor
Resolution (Å)	25.00 – 2.30 24.99 – 2.30	Depositor EDS
% Data completeness (in resolution range)	93.9 (25.00-2.30) 94.0 (24.99-2.30)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	2.58 (at 2.31Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, R_{free}	0.205 , 0.264 0.204 , 0.261	Depositor DCC
R_{free} test set	2705 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å^2)	33.6	Xtriage
Anisotropy	0.619	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 45.6	EDS
L-test for twinning[2]	$<|L|> = 0.50, <L^2> = 0.34$	Xtriage
Estimated twinning fraction	0.010 for -h,-k,l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	9170	wwPDB-VP
Average B, all atoms (Å^2)	51.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.26% 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	0.78	0/2342	0.83	4/3161 (0.1%)
1	D	0.64	0/2319	0.77	1/3131 (0.0%)
2	B	0.99	2/1774 (0.1%)	1.03	8/2393 (0.3%)
2	E	0.80	0/1774	0.86	2/2393 (0.1%)
3	C	0.81	0/398	0.98	4/536 (0.7%)
3	F	0.67	0/398	0.84	0/536
All	All	0.79	2/9005 (0.0%)	0.87	19/12150 (0.2%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	422	SER	CB-OG	6.09	1.50	1.42
2	B	392	GLU	CD-OE1	5.37	1.31	1.25

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	310	ARG	NE-CZ-NH1	7.45	124.03	120.30
2	B	555	MET	CG-SD-CE	-7.37	88.40	100.20
3	C	277	ASP	CB-CG-OD1	-6.26	112.66	118.30
3	C	277	ASP	CB-CG-OD2	6.22	123.90	118.30
2	E	572	ARG	NE-CZ-NH2	-6.14	117.23	120.30

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	2287	0	2341	27	0
1	D	2265	0	2315	25	0
2	B	1737	0	1728	17	0
2	E	1737	0	1728	26	0
3	C	392	0	359	15	0
3	F	392	0	359	13	0
4	A	94	0	0	3	0
4	B	128	0	0	5	0
4	C	17	0	0	0	0
4	D	54	0	0	2	0
4	E	56	0	0	3	0
4	F	11	0	0	0	0
All	All	9170	0	8830	112	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 112 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)
2:B:368:MET:HG3	4:B:688:HOH:O	1.45	1.14
3:F:318:GLU:HA	3:F:318:GLU:OE2	1.58	1.03
3:C:316:GLU:HG3	3:C:316:GLU:O	1.75	0.85
3:C:287:GLU:HG2	3:C:288:PHE:CE2	2.17	0.80
1:D:23:LEU:N	1:D:23:LEU:CD2	2.44	0.79

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	280/297 (94%)	269 (96%)	10 (4%)	1 (0%)	34	42

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	D	277/297 (93%)	266 (96%)	11 (4%)	0	100	100
2	B	209/237 (88%)	203 (97%)	6 (3%)	0	100	100
2	E	209/237 (88%)	204 (98%)	5 (2%)	0	100	100
3	C	47/58 (81%)	46 (98%)	1 (2%)	0	100	100
3	F	47/58 (81%)	44 (94%)	3 (6%)	0	100	100
All	All	1069/1184 (90%)	1032 (96%)	36 (3%)	1 (0%)	51	64

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	262	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	255/266 (96%)	243 (95%)	12 (5%)	26	37
1	D	252/266 (95%)	235 (93%)	17 (7%)	16	21
2	B	192/210 (91%)	184 (96%)	8 (4%)	30	42
2	E	192/210 (91%)	180 (94%)	12 (6%)	18	24
3	C	42/49 (86%)	38 (90%)	4 (10%)	8	10
3	F	42/49 (86%)	40 (95%)	2 (5%)	25	36
All	All	975/1050 (93%)	920 (94%)	55 (6%)	21	29

5 of 55 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	20	SER
1	D	88	ILE
2	E	569	SER
1	D	23	LEU
1	D	50	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	128	HIS
1	D	79	HIS
2	E	504	HIS

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 carbohydrates 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 (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	284/297 (95%)	0.40	28 (9%)	7	10	22, 51, 92, 105	0
1	D	281/297 (94%)	0.74	48 (17%)	1	2	36, 63, 104, 133	0
2	B	213/237 (89%)	0.11	11 (5%)	27	34	18, 31, 62, 82	0
2	E	213/237 (89%)	0.23	12 (5%)	24	30	25, 42, 77, 109	0
3	C	49/58 (84%)	0.78	9 (18%)	1	1	30, 51, 95, 114	0
3	F	49/58 (84%)	1.15	11 (22%)	0	1	30, 61, 127, 159	0
All	All	1089/1184 (91%)	0.45	119 (10%)	5	8	18, 48, 94, 159	0

The worst 5 of 119 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	78	PRO	7.0
3	C	319	SER	5.9
3	F	288	PHE	5.8
3	F	318	GLU	5.6
3	C	288	PHE	5.6

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.