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

Full wwPDB X-ray Structure Validation Report (i)

Nov 18, 2021 – 06:03 PM EST

PDB ID	:	7MZY
Title	:	Anaplastic lymphoma kinase (ALK) extracellular fragment of ligand binding region 673-986
Authors	:	Reshetnyak, A.V.; Sowaileh, M.; Miller, D.J.; Rossi, P.; Myasnikov, A.G.; Kalodimos, C.G.
Deposited on	:	2021-05-24
Resolution	:	1.50 Å(reported)

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

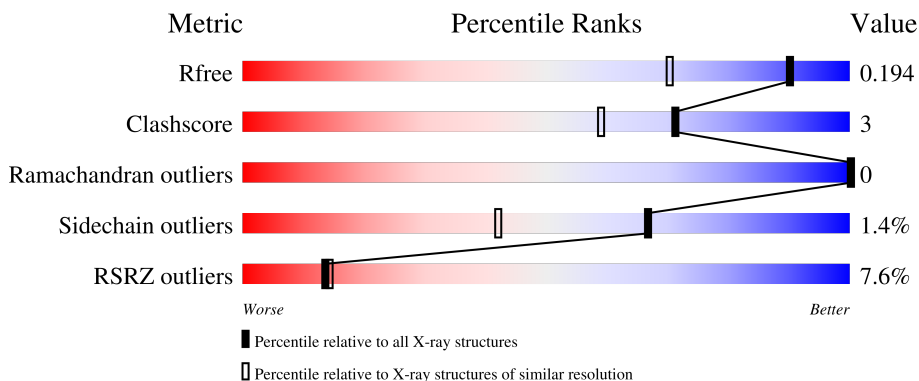|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 1.50 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	Percentile Ranks	Value
Rfree		0.194
Clashscore		3
Ramachandran outliers		0
Sidechain outliers		1.4%
RSRZ outliers		7.6%

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}	130704	2936 (1.50-1.50)
Clashscore	141614	3144 (1.50-1.50)
Ramachandran outliers	138981	3066 (1.50-1.50)
Sidechain outliers	138945	3064 (1.50-1.50)
RSRZ outliers	127900	2884 (1.50-1.50)

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% 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	315	9% / 89% / 9% ·
1	B	315	6% / 90% / 6% ·

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 4868 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 ALK tyrosine kinase receptor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	307	Total	C	N	O	S	5	6	0
			2229	1385	389	442	13			
1	B	302	Total	C	N	O	S	0	5	0
			2179	1347	388	430	14			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	672	GLY	-	expression tag	UNP Q9UM73
B	672	GLY	-	expression tag	UNP Q9UM73

- Molecule 2 is ACETATE ION (three-letter code: ACT) (formula: $C_2H_3O_2$).

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			4	2	2		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total 4	C 2	O 2	0	0

- Molecule 3 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	216	Total 216	O 216	0	0
3	B	236	Total 236	O 236	0	0

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 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: ALK tyrosine kinase receptor
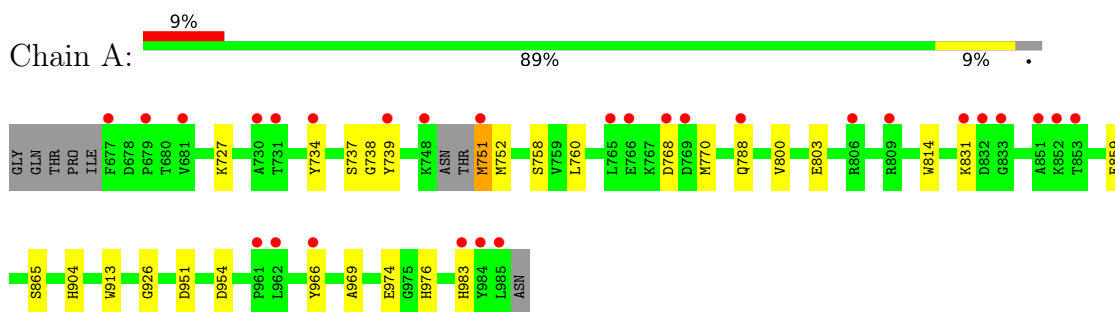
- Molecule 1: ALK tyrosine kinase receptor
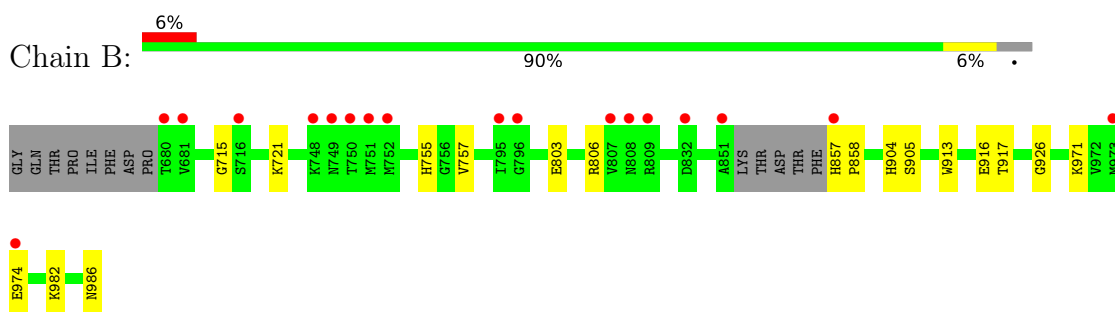
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	70.29Å 133.36Å 62.91Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	37.57 – 1.50 37.57 – 1.50	Depositor EDS
% Data completeness (in resolution range)	80.3 (37.57-1.50) 79.9 (37.57-1.50)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.18 (at 1.50Å)	Xtriage
Refinement program	PHENIX 1.19.1_4122	Depositor
R, R_{free}	0.168 , 0.197 0.166 , 0.194	Depositor DCC
R_{free} test set	5374 reflections (6.75%)	wwPDB-VP
Wilson B-factor (Å^2)	13.0	Xtriage
Anisotropy	0.156	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 54.8	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	4868	wwPDB-VP
Average B, all atoms (Å^2)	22.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.98% 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)

Bond lengths and bond angles in the following residue types are not validated in this section: ACT

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	0.56	0/2290	0.72	1/3097 (0.0%)
1	B	0.60	0/2240	0.75	0/3024
All	All	0.58	0/4530	0.73	1/6121 (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	951	ASP	CB-CG-OD1	5.09	122.88	118.30

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	2229	0	2103	19	0
1	B	2179	0	2069	12	0
2	A	8	0	6	0	0
3	A	216	0	0	3	1
3	B	236	0	0	6	1
All	All	4868	0	4178	30	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 3.

All (30) 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:755:HIS:CE1	1:B:974:GLU:HG3	2.19	0.77
1:A:760:LEU:HD12	1:A:966:TYR:HB3	1.74	0.69
1:B:803:GLU:OE1	1:B:806:ARG:NH2	2.17	0.68
1:A:727:LYS:NZ	1:A:768:ASP:OD1	2.31	0.64
1:A:737:SER:HB2	1:A:760:LEU:HD23	1.83	0.60
1:A:739:TYR:HD2	1:A:974:GLU:OE2	1.89	0.55
1:A:751:MET:HE3	3:A:1275:HOH:O	2.08	0.54
1:B:916:GLU:HG3	3:B:1002:HOH:O	2.08	0.54
1:B:755:HIS:HD2	3:B:1179:HOH:O	1.91	0.53
1:B:917[B]:THR:HG22	3:B:1072:HOH:O	2.07	0.53
1:A:788:GLN:HG2	3:A:1302:HOH:O	2.12	0.49
1:A:734:TYR:CD2	1:A:983:HIS:HB3	2.48	0.49
1:A:803:GLU:OE1	3:A:1101:HOH:O	2.20	0.49
1:A:727:LYS:HG3	1:A:770[A]:MET:SD	2.52	0.49
1:A:760:LEU:HD12	1:A:966:TYR:CB	2.42	0.48
1:A:727:LYS:NZ	1:A:768:ASP:HA	2.30	0.47
1:A:758:SER:O	1:A:969:ALA:HA	2.15	0.46
1:B:757:VAL:HG22	1:B:971:LYS:HG3	1.98	0.46
1:A:859:GLU:OE2	1:A:954:ASP:OD2	2.34	0.45
1:B:715:GLY:N	3:B:1012:HOH:O	2.48	0.45
1:A:738:GLY:O	1:A:758:SER:HA	2.18	0.44
1:A:974:GLU:HG2	1:A:976:HIS:O	2.20	0.42
1:B:982:LYS:NZ	3:B:1014:HOH:O	2.51	0.42
1:B:721:LYS:HE3	3:B:1132:HOH:O	2.19	0.41
1:B:857:HIS:HA	1:B:858:PRO:HD3	1.97	0.41
1:A:800:VAL:HG21	1:A:814:TRP:CE2	2.55	0.41
1:B:904:HIS:O	1:B:926:GLY:HA3	2.20	0.41
1:A:727:LYS:HZ1	1:A:768:ASP:HA	1.86	0.41
1:A:752:MET:HE2	1:B:986:ASN:H	1.85	0.41
1:A:904:HIS:O	1:A:926:GLY:HA3	2.20	0.41

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)
3:A:1261:HOH:O	3:B:1208:HOH:O[4_456]	2.13	0.07

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	309/315 (98%)	303 (98%)	6 (2%)	0	100	100
1	B	303/315 (96%)	299 (99%)	4 (1%)	0	100	100
All	All	612/630 (97%)	602 (98%)	10 (2%)	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	224/228 (98%)	220 (98%)	4 (2%)	59	30
1	B	219/228 (96%)	217 (99%)	2 (1%)	78	61
All	All	443/456 (97%)	437 (99%)	6 (1%)	67	42

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	751	MET
1	A	831	LYS
1	A	865	SER
1	A	913	TRP
1	B	905	SER
1	B	913	TRP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such

sidechains are listed below:

Mol	Chain	Res	Type
1	B	694	HIS
1	B	706	GLN
1	B	755	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 monosaccharides in this entry.

5.6 Ligand geometry ⓘ

2 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	#$|Z| > 2$	Counts	RMSZ	#$|Z| > 2$
2	ACT	A	1002	-	1,3,3	7.44	1 (100%)	0,3,3	-	-
2	ACT	A	1001	-	1,3,3	3.98	1 (100%)	0,3,3	-	-

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1002	ACT	CH3-C	7.44	1.58	1.48
2	A	1001	ACT	CH3-C	3.98	1.53	1.48

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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 ⓘ

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	307/315 (97%)	0.49	28 (9%)	9	9	6, 17, 47, 66	0
1	B	302/315 (95%)	0.30	18 (5%)	21	23	7, 15, 43, 67	0
All	All	609/630 (96%)	0.40	46 (7%)	13	14	6, 16, 46, 67	0

All (46) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	750	THR	11.0
1	A	984	TYR	10.1
1	A	679	PRO	5.8
1	A	966	TYR	5.8
1	B	680	THR	5.7
1	B	749	ASN	5.4
1	A	851	ALA	5.3
1	B	807	VAL	5.2
1	A	751	MET	4.9
1	A	832	ASP	4.4
1	B	796	GLY	4.2
1	A	731	THR	4.0
1	A	831	LYS	4.0
1	B	751	MET	4.0
1	A	853	THR	3.9
1	B	748	LYS	3.8
1	A	769	ASP	3.7
1	A	833	GLY	3.6
1	A	765	LEU	3.6
1	A	809	ARG	3.5
1	A	962	LEU	3.3
1	A	983	HIS	3.2
1	B	851	ALA	3.2
1	A	748	LYS	3.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	974	GLU	3.1
1	B	808	ASN	3.0
1	A	985	LEU	3.0
1	A	852	LYS	3.0
1	B	857	HIS	3.0
1	A	730	ALA	2.9
1	A	734	TYR	2.9
1	B	832	ASP	2.8
1	B	752	MET	2.7
1	B	681	VAL	2.6
1	B	809	ARG	2.6
1	A	766	GLU	2.5
1	A	677	PHE	2.4
1	B	795	ILE	2.4
1	A	788	GLN	2.3
1	A	681	VAL	2.3
1	A	806	ARG	2.3
1	A	768	ASP	2.2
1	B	716	SER	2.1
1	B	973	MET	2.1
1	A	961	PRO	2.1
1	A	739	TYR	2.0

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

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors($Å^2$)	Q<0.9
2	ACT	A	1001	4/4	0.88	0.16	34,39,45,51	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors($\mathring{A}^2$)	Q<0.9
2	ACT	A	1002	4/4	0.89	0.14	20,25,38,38	0

6.5 Other polymers ⓘ

There are no such residues in this entry.