WORLDWIDE PDB
PROTEIN DATA BANK

Full wwPDB X-ray Structure Validation Report ⓘ

Oct 15, 2023 – 10:06 AM EDT

PDB ID	:	7TVD
Title	:	Crystal structure of the kinase domain of EGFR exon-19 (del-747-749) mutant
Authors	:	Ashtekar, K.D.; Stayrook, S.E.; Lemmon, M.A.
Deposited on	:	2022-02-04
Resolution	:	2.96 Å(reported)

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

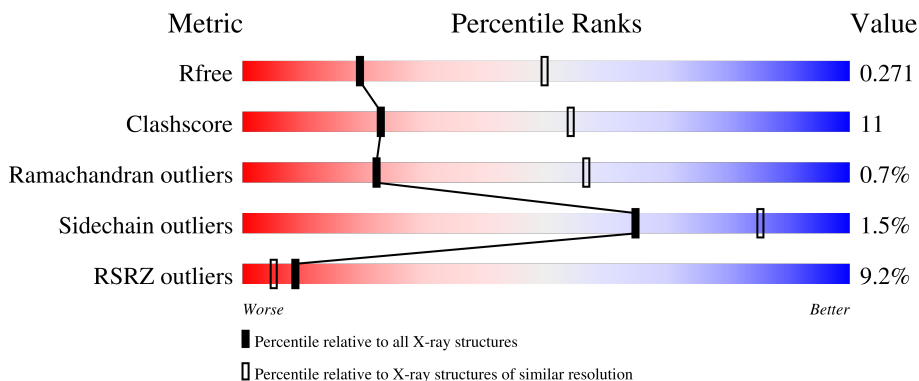|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.96 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		0.271
Clashscore		11
Ramachandran outliers		0.7%
Sidechain outliers		1.5%
RSRZ outliers		9.2%

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	3104 (3.00-2.92)
Clashscore	141614	3462 (3.00-2.92)
Ramachandran outliers	138981	3340 (3.00-2.92)
Sidechain outliers	138945	3343 (3.00-2.92)
RSRZ outliers	127900	2986 (3.00-2.92)

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	335	8% · 62% · 25% · 13%

2 Entry composition ⓘ

There is only 1 type of molecule in this entry. The entry contains 2338 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 Epidermal growth factor receptor.

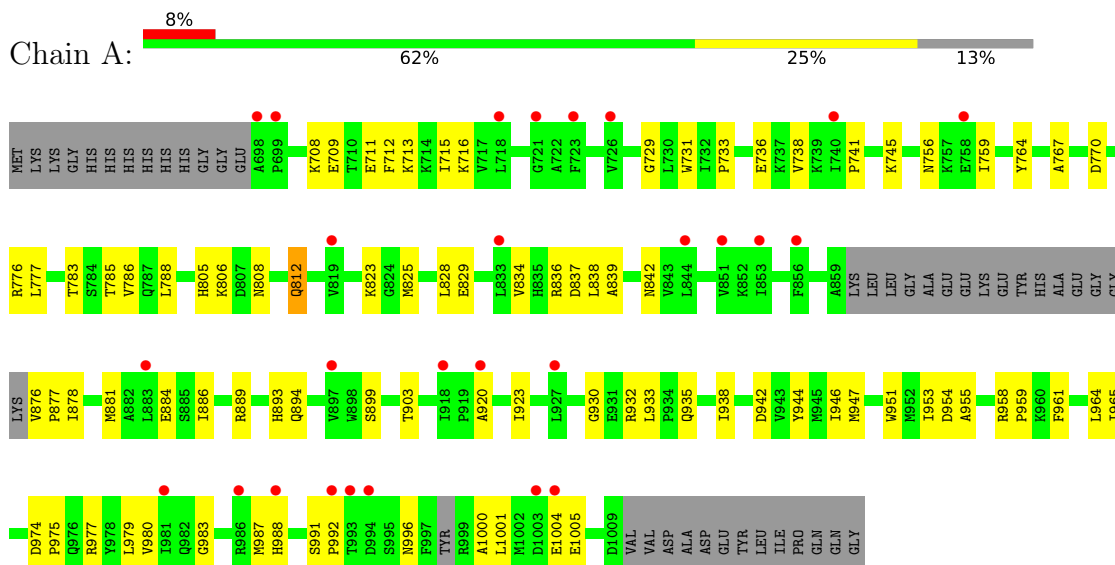| Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	292	Total 2338	C 1500	N 395	O 425	S 18	0	0	0

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	685	MET	-	initiating methionine	UNP P00533
A	686	LYS	-	expression tag	UNP P00533
A	687	LYS	-	expression tag	UNP P00533
A	688	GLY	-	expression tag	UNP P00533
A	689	HIS	-	expression tag	UNP P00533
A	690	HIS	-	expression tag	UNP P00533
A	691	HIS	-	expression tag	UNP P00533
A	692	HIS	-	expression tag	UNP P00533
A	693	HIS	-	expression tag	UNP P00533
A	694	HIS	-	expression tag	UNP P00533
A	695	GLY	-	expression tag	UNP P00533
A	?	-	LEU	deletion	UNP P00533
A	?	-	ARG	deletion	UNP P00533
A	?	-	GLU	deletion	UNP P00533

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: Epidermal growth factor receptor

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	I 2 3	Depositor
Cell constants a, b, c, α, β, γ	149.42Å 149.42Å 149.42Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	74.71 – 2.96 105.66 – 2.96	Depositor EDS
% Data completeness (in resolution range)	99.7 (74.71-2.96) 99.8 (105.66-2.96)	Depositor EDS
R_{merge}	0.26	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.48 (at 2.96Å)	Xtriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, R_{free}	0.240 , 0.277 0.236 , 0.271	Depositor DCC
R_{free} test set	562 reflections (4.79%)	wwPDB-VP
Wilson B-factor (Å^2)	109.5	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 76.6	EDS
L-test for twinning [2]	$<|L|> = 0.50$, $<L^2> = 0.33$	Xtriage
Estimated twinning fraction	0.008 for -l,-k,-h	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	2338	wwPDB-VP
Average B, all atoms (Å^2)	104.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.11% 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	0.28	0/2389	0.55	0/3234

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	2338	0	2371	54	1
All	All	2338	0	2371	54	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 11.

All (54) 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:709:GLU:H	1:A:709:GLU:CD	1.90	0.75
1:A:745:LYS:HZ2	1:A:759:ILE:HD11	1.64	0.63
1:A:770:ASP:HA	1:A:776:ARG:HH12	1.64	0.63
1:A:812:GLN:HG2	1:A:975:PRO:HG2	1.81	0.61
1:A:770:ASP:OD1	1:A:776:ARG:NH1	2.33	0.61

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:834:VAL:HG22	1:A:893:HIS:CD2	2.37	0.60
1:A:961:PHE:O	1:A:965:ILE:HG13	2.02	0.59
1:A:834:VAL:HG22	1:A:893:HIS:HD2	1.66	0.59
1:A:953:ILE:HD12	1:A:954:ASP:N	2.17	0.59
1:A:708:LYS:O	1:A:711:GLU:HB2	2.04	0.58
1:A:783:THR:OG1	1:A:785:THR:O	2.20	0.58
1:A:745:LYS:HB3	1:A:788:LEU:HB3	1.86	0.58
1:A:741:PRO:HD2	1:A:1004:GLU:O	2.06	0.56
1:A:899:SER:O	1:A:903:THR:HG23	2.07	0.54
1:A:946:ILE:HD12	1:A:947:MET:N	2.22	0.54
1:A:996:ASN:HD22	1:A:1000:ALA:HB3	1.73	0.53
1:A:1004:GLU:N	1:A:1004:GLU:OE1	2.45	0.50
1:A:1005:GLU:OE1	1:A:1005:GLU:N	2.45	0.50
1:A:959:PRO:HB2	1:A:964:LEU:HD21	1.94	0.49
1:A:980:VAL:HG23	1:A:980:VAL:O	2.12	0.49
1:A:715:ILE:HG12	1:A:729:GLY:HA2	1.93	0.49
1:A:731:TRP:CH2	1:A:733:PRO:HG3	2.48	0.49
1:A:825:MET:SD	1:A:838:LEU:HD22	2.54	0.48
1:A:756:ASN:O	1:A:759:ILE:N	2.47	0.48
1:A:983:GLY:O	1:A:987:MET:HG3	2.14	0.48
1:A:715:ILE:HG22	1:A:716:LYS:HG3	1.95	0.48
1:A:834:VAL:HG12	1:A:836:ARG:HG3	1.97	0.47
1:A:837:ASP:O	1:A:842:ASN:ND2	2.46	0.47
1:A:825:MET:HE1	1:A:828:LEU:HD22	1.97	0.46
1:A:894:GLN:HB2	1:A:955:ALA:HB1	1.97	0.46
1:A:932:ARG:HE	1:A:951:TRP:HB3	1.81	0.46
1:A:974:ASP:O	1:A:977:ARG:HB3	2.15	0.46
1:A:877:PRO:O	1:A:881:MET:HG3	2.17	0.45
1:A:709:GLU:HA	1:A:712:PHE:CE2	2.52	0.44
1:A:829:GLU:HA	1:A:893:HIS:CE1	2.53	0.44
1:A:938:ILE:HD12	1:A:979:LEU:HD22	2.00	0.44
1:A:935:GLN:HA	1:A:944:TYR:CE2	2.53	0.44
1:A:805:HIS:O	1:A:808:ASN:N	2.50	0.43
1:A:733:PRO:HB2	1:A:736:GLU:HB2	2.00	0.43
1:A:920:ALA:HA	1:A:923:ILE:HD12	2.01	0.43
1:A:884:GLU:OE2	1:A:958:ARG:NH1	2.45	0.43
1:A:759:ILE:HG21	1:A:786:VAL:HG23	2.01	0.43
1:A:1001:LEU:HD23	1:A:1001:LEU:HA	1.89	0.43
1:A:709:GLU:CD	1:A:709:GLU:N	2.65	0.43
1:A:839:ALA:HA	1:A:903:THR:HG22	2.01	0.42
1:A:731:TRP:CZ3	1:A:733:PRO:HG3	2.54	0.41

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:991:SER:OG	1:A:992:PRO:HD2	2.20	0.41
1:A:889:ARG:HG3	1:A:889:ARG:HH11	1.86	0.41
1:A:767:ALA:HB2	1:A:777:LEU:HD23	2.03	0.41
1:A:823:LYS:HG2	1:A:965:ILE:HD13	2.02	0.41
1:A:878:ILE:HD11	1:A:886:ILE:HD13	2.03	0.41
1:A:825:MET:CE	1:A:825:MET:HA	2.51	0.40
1:A:738:VAL:HG13	1:A:1004:GLU:CD	2.42	0.40
1:A:933:LEU:O	1:A:944:TYR:OH	2.37	0.40

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)
1:A:764:TYR:OH	1:A:942:ASP:OD1[18_445]	2.19	0.01

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	286/335 (85%)	269 (94%)	15 (5%)	2 (1%)	22	56

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	806	LYS
1	A	930	GLY

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	260/293 (89%)	256 (98%)	4 (2%)	65 85

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	713	LYS
1	A	812	GLN
1	A	876	VAL
1	A	988	HIS

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	996	ASN

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 (i)

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	292/335 (87%)	0.91	27 (9%)	9	5	75, 97, 154, 184	0

All (27) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	721	GLY	4.8
1	A	988	HIS	4.0
1	A	856	PHE	3.6
1	A	994	ASP	3.4
1	A	1004	GLU	3.3
1	A	833	LEU	3.0
1	A	699	PRO	2.8
1	A	844	LEU	2.7
1	A	1003	ASP	2.7
1	A	986	ARG	2.6
1	A	927	LEU	2.4
1	A	853	ILE	2.4
1	A	883	LEU	2.3
1	A	918	ILE	2.3
1	A	992	PRO	2.3
1	A	740	ILE	2.2
1	A	819	VAL	2.2
1	A	718	LEU	2.2
1	A	981	ILE	2.2
1	A	851	VAL	2.1
1	A	897	VAL	2.1
1	A	920	ALA	2.1
1	A	726	VAL	2.1
1	A	993	THR	2.1
1	A	723	PHE	2.0
1	A	698	ALA	2.0
1	A	758	GLU	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 ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.