



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 22, 2016 – 07:56 PM EDT

PDB ID : 4XEV  
Title : Fusion of Pyk2-FAT domain with Leupaxin LD1 motif, complexed with Leupaxin LD4 peptide  
Authors : Miller, D.J.  
Deposited on : 2014-12-24  
Resolution : 2.01 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](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity : 4.02b-467  
Mogul : unknown  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20027107  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0122  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20027107

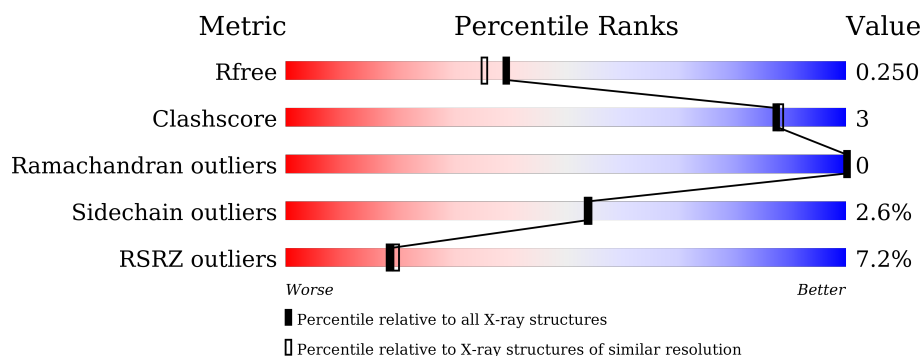
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.01 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))
$R_{free}$	91344	6249 (2.00-2.00)
Clashscore	102246	7340 (2.00-2.00)
Ramachandran outliers	100387	7248 (2.00-2.00)
Sidechain outliers	100360	7247 (2.00-2.00)
RSRZ outliers	91569	6262 (2.00-2.00)

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  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. 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  $\leq 5\%$ . 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	167	<div> <div>6%</div> <div> <div></div> <div>81%</div> <div>7%</div> <div>11%</div> </div> </div>
1	D	167	<div> <div>3%</div> <div> <div></div> <div>83%</div> <div>5%</div> <div>12%</div> </div> </div>
2	C	19	<div> <div>26%</div> <div> <div></div> <div>89%</div> <div>11%</div> </div> </div>
2	F	19	<div> <div>21%</div> <div> <div></div> <div>79%</div> <div>16%</div> <div>5%</div> </div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 2579 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 Fusion protein of Protein-tyrosine kinase 2-beta FAT domain and Leupaxin LD1 motif.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	148	Total	C	N	O	S	0	0	0
			1111	696	193	219	3			
1	D	147	Total	C	N	O	S	0	4	0
			1136	715	190	226	5			

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	867	GLY	-	expression tag	UNP Q14289
A	868	SER	-	expression tag	UNP Q14289
A	869	HIS	-	expression tag	UNP Q14289
A	870	MET	-	expression tag	UNP Q14289
A	899	SER	CYS	engineered mutation	UNP Q14289
A	972	ALA	CYS	engineered mutation	UNP Q14289
A	1006	GLY	-	linker	UNP Q14289
A	1007	GLY	-	linker	UNP Q14289
A	1995	SER	-	linker	UNP Q14289
A	1996	GLY	-	linker	UNP Q14289
A	1997	GLY	-	linker	UNP Q14289
A	1998	GLY	-	linker	UNP Q14289
A	1999	GLY	-	linker	UNP Q14289
A	2000	GLY	-	linker	UNP Q14289
D	867	GLY	-	expression tag	UNP Q14289
D	868	SER	-	expression tag	UNP Q14289
D	869	HIS	-	expression tag	UNP Q14289
D	870	MET	-	expression tag	UNP Q14289
D	899	SER	CYS	engineered mutation	UNP Q14289
D	972	ALA	CYS	engineered mutation	UNP Q14289
D	1006	GLY	-	linker	UNP Q14289
D	1994	GLY	-	linker	UNP Q14289
D	1995	SER	-	linker	UNP Q14289
D	1996	GLY	-	linker	UNP Q14289

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
D	1997	GLY	-	linker	UNP Q14289
D	1998	GLY	-	linker	UNP Q14289
D	1999	GLY	-	linker	UNP Q14289
D	2000	GLY	-	linker	UNP Q14289

- Molecule 2 is a protein called 19-mer peptide containing Leupaxin LD4 motif.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	19	Total 135	C 82	N 22	O 29	S 2	0	0	0
2	F	18	Total 126	C 78	N 22	O 24	S 2	0	0	0

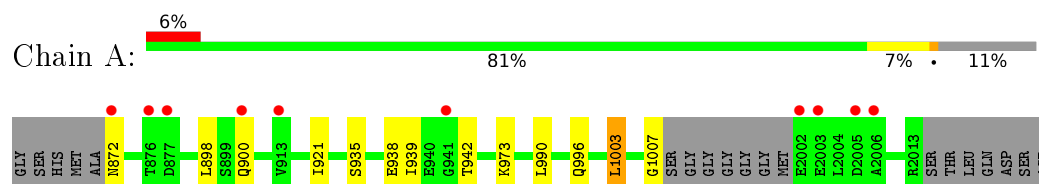
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	32	Total 32	O 32	0	0
3	D	38	Total 38	O 38	0	0
3	F	1	Total 1	O 1	0	0

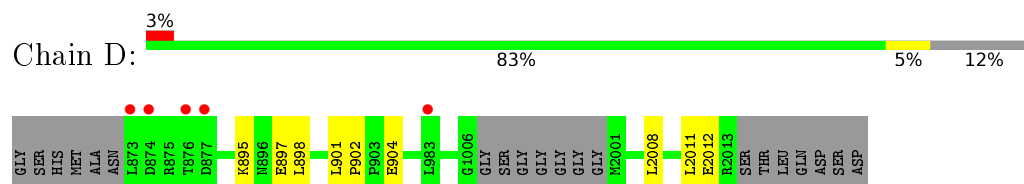
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

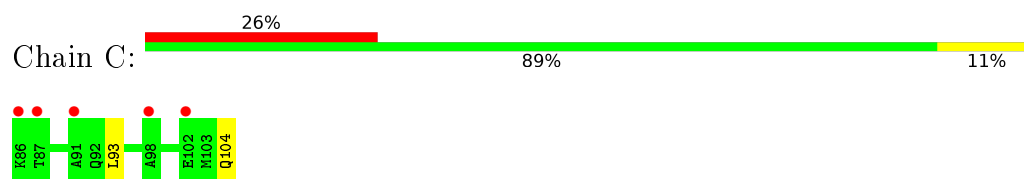
- Molecule 1: Fusion protein of Protein-tyrosine kinase 2-beta FAT domain and Leupaxin LD1 motif



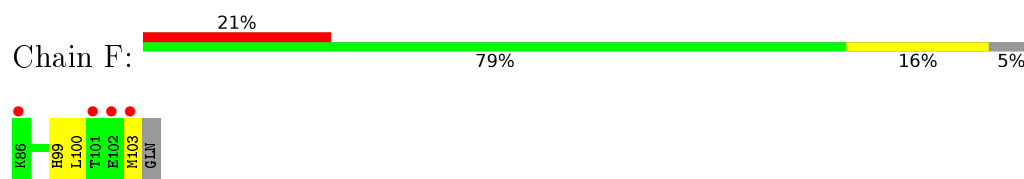
- Molecule 1: Fusion protein of Protein-tyrosine kinase 2-beta FAT domain and Leupaxin LD1 motif



- Molecule 2: 19-mer peptide containing Leupaxin LD4 motif



- Molecule 2: 19-mer peptide containing Leupaxin LD4 motif



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	52.08 Å 79.23 Å 53.23 Å 90.00° 117.62° 90.00°	Depositor
Resolution (Å)	27.26 – 2.01 27.26 – 2.01	Depositor EDS
% Data completeness (in resolution range)	95.8 (27.26-2.01) 95.9 (27.26-2.01)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.04	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.77 (at 2.01 Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, $R_{free}$	0.204 , 0.250 0.204 , 0.250	Depositor DCC
$R_{free}$ test set	1275 reflections (5.19%)	DCC
Wilson B-factor (Å <sup>2</sup> )	41.5	Xtriage
Anisotropy	0.659	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 54.7	EDS
Estimated twinning fraction	0.016 for l,-k,h	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.34$	Xtriage
Outliers	0 of 24597 reflections	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	2579	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	60.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.06% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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	0.35	0/1117	0.48	0/1514
1	D	0.39	0/1153	0.53	0/1560
2	C	0.22	0/135	0.41	0/182
2	F	0.26	0/126	0.45	0/169
All	All	0.36	0/2531	0.50	0/3425

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	1111	0	1142	6	0
1	D	1136	0	1192	5	0
2	C	135	0	124	1	0
2	F	126	0	125	2	0
3	A	32	0	0	1	0
3	D	38	0	0	0	0
3	F	1	0	0	0	0
All	All	2579	0	2583	13	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 3.

The worst 5 of 13 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:921:ILE:HD12	2:C:93:LEU:HD22	1.80	0.64
1:A:942:THR:OG1	1:A:990:LEU:HD13	1.99	0.63
1:D:2008:LEU:O	1:D:2012:GLU:HG2	2.05	0.57
1:A:935:SER:O	1:A:939:ILE:HG12	2.08	0.53
1:D:897:GLU:O	1:D:901:LEU:HG	2.12	0.49

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	144/167 (86%)	143 (99%)	1 (1%)	0	100	100
1	D	147/167 (88%)	147 (100%)	0	0	100	100
2	C	17/19 (90%)	16 (94%)	1 (6%)	0	100	100
2	F	16/19 (84%)	16 (100%)	0	0	100	100
All	All	324/372 (87%)	322 (99%)	2 (1%)	0	100	100

There are no Ramachandran outliers to report.

### 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	121/141 (86%)	116 (96%)	5 (4%)	37	32
1	D	129/141 (92%)	128 (99%)	1 (1%)	86	89
2	C	13/15 (87%)	12 (92%)	1 (8%)	16	10
2	F	12/15 (80%)	12 (100%)	0	100	100
All	All	275/312 (88%)	268 (98%)	7 (2%)	54	55

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	973	LYS
2	C	104	GLN
1	A	1003	LEU
1	A	900	GLN
1	D	898	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	996	GLN
1	D	900	GLN

### 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 carbohydrates 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](#)

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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	148/167 (88%)	0.38	10 (6%) 20 22	38, 58, 101, 119	0
1	D	147/167 (88%)	0.19	5 (3%) 49 50	36, 53, 88, 102	0
2	C	19/19 (100%)	1.25	5 (26%) 1 1	59, 75, 100, 104	0
2	F	18/19 (94%)	0.86	4 (22%) 1 1	59, 71, 98, 107	0
All	All	332/372 (89%)	0.37	24 (7%) 18 20	36, 57, 97, 119	0

The worst 5 of 24 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	876	THR	4.9
2	F	86	LYS	3.4
2	F	101	THR	3.4
1	D	876	THR	3.3
1	A	2006	ALA	3.3

### 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.