



# Full wwPDB X-ray Structure Validation Report ⓘ

Mar 21, 2016 – 01:01 PM EDT

PDB ID : 4XKH  
Title : CRYSTAL STRUCTURE OF THE AIRAPL TANDEM UIMS IN COMPLEX  
WITH A LYS48-LINKED TRI-UBIQUITIN  
Authors : Rahighi, S.; Kawasaki, M.; Stanhill, A.; Wakatsuki, S.  
Deposited on : 2015-01-11  
Resolution : 3.00 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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.

---

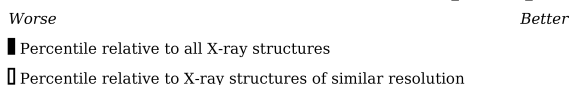
The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

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

**i**

## X-RAY DIFFRACTION

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.



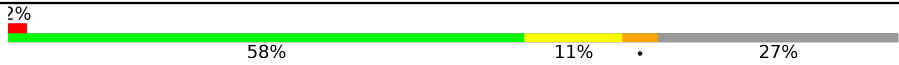


Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	91344	1578 (3.00-3.00)
Clashscore	102246	1912 (3.00-3.00)
Ramachandran outliers	100387	1853 (3.00-3.00)
Sidechain outliers	100360	1856 (3.00-3.00)
RSRZ outliers	91569	1592 (3.00-3.00)

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	76	<div><div></div><div></div><div></div></div> 87%7%5%
1	B	76	<div><div></div><div></div><div></div></div> 84%12%..
1	D	76	<div><div></div><div></div><div></div></div> 92%...
1	F	76	<div><div></div><div></div><div></div></div> 87%9%.
1	G	76	<div><div></div><div></div><div></div></div> 92%8%
1	I	76	<div><div></div><div></div><div></div></div> 88%5%..

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	C	55	
2	E	55	
2	H	55	

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 4421 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 Polyubiquitin-C.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
1	A	72	Total	C	N	O	0	0	0
			574	363	98	113			
1	F	73	Total	C	N	O	0	0	0
			582	368	99	114			
1	B	74	Total	C	N	O	0	0	0
			593	374	103	115			
1	D	73	Total	C	N	O	0	0	0
			582	368	99	114			
1	G	76	Total	C	N	O	0	0	0
			597	375	104	117			
1	I	73	Total	C	N	O	0	0	0
			582	368	99	114			

- Molecule 2 is a protein called AN1-type zinc finger protein 2B.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	E	43	Total	C	N	O	0	0	0
			317	191	52	74			
2	C	40	Total	C	N	O	0	0	0
			283	170	44	69			
2	H	40	Total	C	N	O	0	0	0
			296	179	48	69			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	186	GLY	-	expression tag	UNP Q91X58
C	186	GLY	-	expression tag	UNP Q91X58
H	186	GLY	-	expression tag	UNP Q91X58

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	3	Total 3	O 3	0	0
3	E	2	Total 2	O 2	0	0
3	F	2	Total 2	O 2	0	0
3	B	2	Total 2	O 2	0	0
3	D	1	Total 1	O 1	0	0
3	G	4	Total 4	O 4	0	0
3	H	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: Polyubiquitin-C

Chain A:  87% 7% • 5%




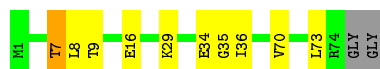
- Molecule 1: Polyubiquitin-C

Chain F:  87% 9% •



- Molecule 1: Polyubiquitin-C

Chain B:  84% 12% • •



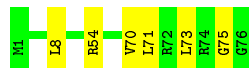
- Molecule 1: Polyubiquitin-C

Chain D:  92% • • •




- Molecule 1: Polyubiquitin-C

Chain G:  92% 8%



- Molecule 1: Polyubiquitin-C

Chain I:  88% 5% • •



- Molecule 2: AN1-type zinc finger protein 2B

Chain E: 60% 15% 22%



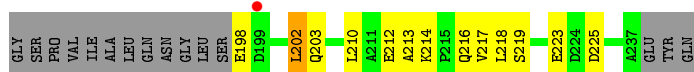
- Molecule 2: AN1-type zinc finger protein 2B

Chain C: 2% 58% 11% 27%



- Molecule 2: AN1-type zinc finger protein 2B

Chain H: 2% 49% 22% 27%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 31	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	90.86 Å 90.86 Å 61.01 Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	48.22 – 3.00 48.22 – 3.00	Depositor EDS
% Data completeness (in resolution range)	99.3 (48.22-3.00) 99.3 (48.22-3.00)	Depositor EDS
$R_{merge}$	0.16	Depositor
$R_{sym}$	0.23	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.17 (at 3.01 Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, $R_{free}$	0.201 , 0.255 0.203 , 0.260	Depositor DCC
$R_{free}$ test set	537 reflections (5.02%)	DCC
Wilson B-factor (Å <sup>2</sup> )	49.9	Xtriage
Anisotropy	0.405	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 6.4	EDS
Estimated twinning fraction	0.615 for H, K, L 0.385 for -K, -H, -L 0.117 for -h,-k,l 0.124 for h,-h-k,-l 0.319 for -k,-h,-l	Xtriage
Reported twinning fraction	0.615 for H, K, L 0.385 for -K, -H, -L	Depositor
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.39$ , $\langle L^2 \rangle = 0.21$	Xtriage
Outliers	0 of 11212 reflections	Xtriage
$F_o, F_c$ correlation	0.88	EDS
Total number of atoms	4421	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	47.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.03% 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	0.52	0/580	0.61	1/782 (0.1%)
1	B	0.63	0/599	0.63	0/806
1	D	0.65	0/588	0.69	1/792 (0.1%)
1	F	0.63	0/588	0.62	0/792
1	G	0.59	0/603	0.62	0/812
1	I	0.55	0/588	0.62	0/792
2	C	0.72	0/283	0.69	0/385
2	E	0.64	0/317	0.63	0/428
2	H	0.63	0/296	0.70	0/400
All	All	0.61	0/4442	0.64	2/5989 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	D	32	ASP	CB-CG-OD1	5.63	123.37	118.30
1	A	58	ASP	CB-CG-OD1	5.49	123.24	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	574	0	598	1	0
1	B	593	0	623	4	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	D	582	0	610	2	0
1	F	582	0	610	4	0
1	G	597	0	618	2	0
1	I	582	0	610	4	0
2	C	283	0	254	4	0
2	E	317	0	305	7	0
2	H	296	0	285	5	0
3	A	3	0	0	0	0
3	B	2	0	0	0	0
3	D	1	0	0	0	0
3	E	2	0	0	0	0
3	F	2	0	0	0	0
3	G	4	0	0	0	0
3	H	1	0	0	0	0
All	All	4421	0	4513	29	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.

All (29) 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:F:73:LEU:O	1:F:73:LEU:HD13	1.48	1.13
1:F:73:LEU:O	1:F:73:LEU:CD1	2.25	0.85
1:F:73:LEU:HD13	1:F:73:LEU:C	2.02	0.80
1:D:73:LEU:HD13	1:D:73:LEU:C	2.17	0.64
2:E:217:VAL:HG12	2:E:217:VAL:O	2.04	0.57
2:H:213:ALA:O	2:H:219:SER:OG	2.23	0.56
1:D:73:LEU:CD1	1:D:73:LEU:C	2.75	0.54
1:I:73:LEU:C	1:I:73:LEU:HD13	2.28	0.54
2:H:217:VAL:O	2:H:217:VAL:HG12	2.08	0.54
1:A:7:THR:HB	2:E:215:PRO:HG3	1.91	0.53
1:G:8:LEU:HD11	2:H:225:ASP:HA	1.91	0.52
2:H:198:GLU:HB2	2:H:202:LEU:HD12	1.95	0.48
2:E:200:GLU:O	2:E:204:ARG:HG2	2.14	0.48
2:C:226:LEU:O	2:C:230:GLN:N	2.43	0.47
2:E:214:LYS:HB3	2:E:215:PRO:HD3	1.96	0.47
1:I:73:LEU:C	1:I:73:LEU:CD1	2.83	0.47
1:I:16:GLU:O	1:I:29:LYS:NZ	2.48	0.47
2:E:204:ARG:O	2:E:208:LEU:HD12	2.15	0.47
1:I:71:LEU:HD12	1:I:71:LEU:N	2.30	0.46

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:H:216:GLN:N	2:H:216:GLN:OE1	2.50	0.44
1:B:34:GLU:O	1:B:36:ILE:N	2.51	0.44
1:B:16:GLU:O	1:B:29:LYS:NZ	2.49	0.44
2:E:222:GLU:O	2:E:226:LEU:HB2	2.18	0.43
2:C:219:SER:OG	2:C:220:SER:N	2.51	0.43
1:G:70:VAL:C	1:G:71:LEU:HD12	2.39	0.42
1:F:24:GLU:HG2	1:F:52:ASP:HB3	2.00	0.42
1:B:7:THR:HG21	2:C:215:PRO:HB3	2.01	0.42
1:B:9:THR:HG22	2:C:215:PRO:HG2	2.03	0.41
2:E:214:LYS:HG2	2:E:215:PRO:N	2.35	0.41

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	70/76 (92%)	65 (93%)	5 (7%)	0	100	100
1	B	72/76 (95%)	67 (93%)	4 (6%)	1 (1%)	14	51
1	D	71/76 (93%)	69 (97%)	2 (3%)	0	100	100
1	F	71/76 (93%)	69 (97%)	2 (3%)	0	100	100
1	G	74/76 (97%)	70 (95%)	3 (4%)	1 (1%)	14	51
1	I	71/76 (93%)	68 (96%)	3 (4%)	0	100	100
2	C	38/55 (69%)	32 (84%)	3 (8%)	3 (8%)	1	5
2	E	41/55 (74%)	32 (78%)	8 (20%)	1 (2%)	7	35
2	H	38/55 (69%)	34 (90%)	3 (8%)	1 (3%)	7	33
All	All	546/621 (88%)	506 (93%)	33 (6%)	7 (1%)	15	53

All (7) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	214	LYS
2	C	215	PRO
2	H	214	LYS
2	E	214	LYS
2	C	219	SER
1	G	75	GLY
1	B	35	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	66/68 (97%)	61 (92%)	5 (8%)	16	51
1	B	68/68 (100%)	64 (94%)	4 (6%)	24	63
1	D	67/68 (98%)	65 (97%)	2 (3%)	48	83
1	F	67/68 (98%)	63 (94%)	4 (6%)	24	62
1	G	67/68 (98%)	65 (97%)	2 (3%)	48	83
1	I	67/68 (98%)	63 (94%)	4 (6%)	24	62
2	C	28/43 (65%)	26 (93%)	2 (7%)	18	54
2	E	34/43 (79%)	31 (91%)	3 (9%)	12	42
2	H	31/43 (72%)	25 (81%)	6 (19%)	2	9
All	All	495/537 (92%)	463 (94%)	32 (6%)	21	58

All (32) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	2	GLN
1	A	7	THR
1	A	16	GLU
1	A	62	GLN
1	A	70	VAL
2	E	218	LEU
2	E	225	ASP
2	E	226	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	F	5	VAL
1	F	33	LYS
1	F	49	GLN
1	F	63	LYS
1	B	7	THR
1	B	8	LEU
1	B	70	VAL
1	B	73	LEU
2	C	210	LEU
2	C	228	LEU
1	D	13	ILE
1	D	73	LEU
1	G	54	ARG
1	G	73	LEU
2	H	202	LEU
2	H	203	GLN
2	H	210	LEU
2	H	212	GLU
2	H	218	LEU
2	H	223	GLU
1	I	29	LYS
1	I	63	LYS
1	I	64	GLU
1	I	73	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	2	GLN
2	E	230	GLN
1	F	49	GLN
1	B	49	GLN
2	H	203	GLN

### 5.3.3 RNA ⓘ

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

### 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, 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	72/76 (94%)	-0.29	0	100 100	34, 49, 61, 69	0
1	B	74/76 (97%)	-0.30	0	100 100	29, 42, 57, 66	0
1	D	73/76 (96%)	-0.37	0	100 100	22, 43, 56, 61	1 (1%)
1	F	73/76 (96%)	-0.42	0	100 100	27, 41, 54, 68	1 (1%)
1	G	76/76 (100%)	-0.23	0	100 100	33, 51, 68, 73	0
1	I	73/76 (96%)	-0.21	0	100 100	33, 51, 64, 71	1 (1%)
2	C	40/55 (72%)	-0.17	1 (2%)	61 30	31, 43, 62, 74	0
2	E	43/55 (78%)	-0.14	0	100 100	30, 48, 65, 67	0
2	H	40/55 (72%)	-0.06	1 (2%)	61 30	35, 51, 66, 72	0
All	All	564/621 (90%)	-0.26	2 (0%)	93 80	22, 47, 64, 74	3 (0%)

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	H	199	ASP	2.5
2	C	216	GLN	2.2

### 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 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.