



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 13, 2017 – 05:28 am GMT

PDB ID : 5L8J  
Title : Aurora-A kinase domain in complex with vNAR-D01 S93R  
Authors : Burgess, S.G.; Bayliss, R.  
Deposited on : 2016-06-08  
Resolution : 1.68 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 : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : trunk28620  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : recalc28949

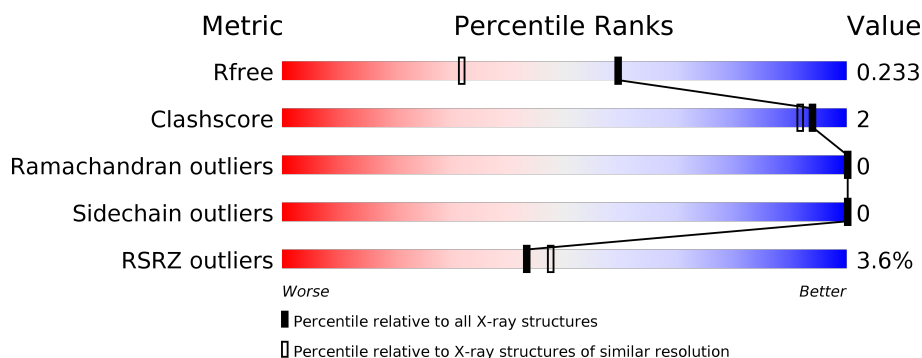
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.68 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))
$R_{free}$	100719	5252 (1.70-1.66)
Clashscore	112137	5803 (1.70-1.66)
Ramachandran outliers	110173	5704 (1.70-1.66)
Sidechain outliers	110143	5703 (1.70-1.66)
RSRZ outliers	101464	5298 (1.70-1.66)

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	285	<div> <div>3%</div> <div> <div></div> <div>86%</div> <div>•</div> <div>11%</div> </div> </div>
2	B	117	<div> <div>4%</div> <div> <div></div> <div>87%</div> <div>•</div> <div>10%</div> </div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
5	PEG	A	505	-	-	-	X

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 3041 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 Aurora kinase A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	253	Total	C	N	O	S	0	2	0
			2024	1313	341	366	4			

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	119	GLY	-	expression tag	UNP O14965
A	120	ALA	-	expression tag	UNP O14965
A	121	MET	-	expression tag	UNP O14965
A	290	ALA	CYS	engineered mutation	UNP O14965
A	393	ALA	CYS	engineered mutation	UNP O14965

- Molecule 2 is a protein called New antigen receptor variable domain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	105	Total	C	N	O	S	0	0	1
			771	467	136	164	4			

There are 26 discrepancies between the modelled and reference sequences:

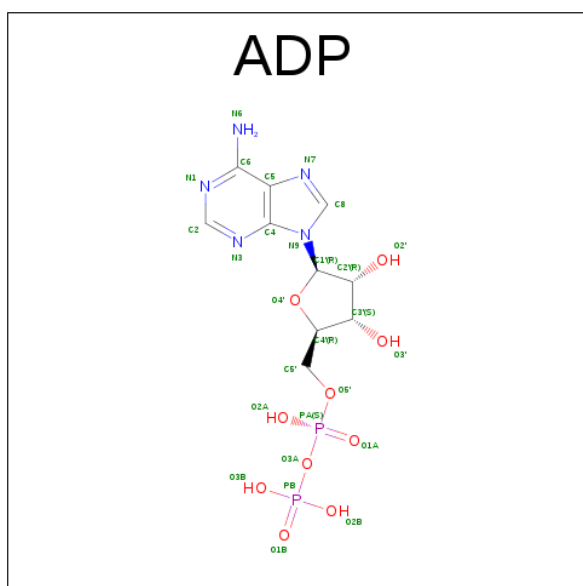
Chain	Residue	Modelled	Actual	Comment	Reference
B	1	MET	-	initiating methionine	UNP Q8JJ25
B	87	ILE	TYR	conflict	UNP Q8JJ25
B	88	ASP	ARG	conflict	UNP Q8JJ25
B	89	SER	ARG	engineered mutation	UNP Q8JJ25
B	?	-	ALA	deletion	UNP Q8JJ25
B	?	-	PHE	deletion	UNP Q8JJ25
B	?	-	ASN	deletion	UNP Q8JJ25
B	?	-	THR	deletion	UNP Q8JJ25
B	?	-	GLY	deletion	UNP Q8JJ25
B	91	TRP	VAL	conflict	UNP Q8JJ25
B	92	LEU	GLY	conflict	UNP Q8JJ25

*Continued on next page...*

Continued from previous page...

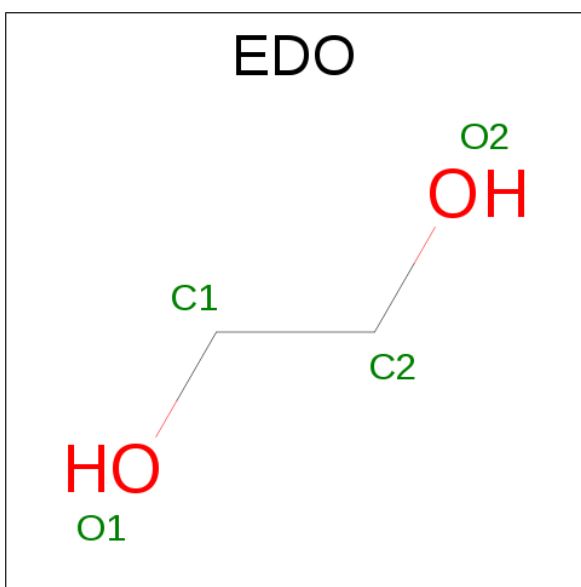
Chain	Residue	Modelled	Actual	Comment	Reference
B	93	ARG	TYR	conflict	UNP Q8JJ25
B	94	ARG	LYS	conflict	UNP Q8JJ25
B	105	GLY	-	expression tag	UNP Q8JJ25
B	106	GLY	-	expression tag	UNP Q8JJ25
B	107	ALA	-	expression tag	UNP Q8JJ25
B	108	ALA	-	expression tag	UNP Q8JJ25
B	109	ALA	-	expression tag	UNP Q8JJ25
B	110	LEU	-	expression tag	UNP Q8JJ25
B	111	GLU	-	expression tag	UNP Q8JJ25
B	112	HIS	-	expression tag	UNP Q8JJ25
B	113	HIS	-	expression tag	UNP Q8JJ25
B	114	HIS	-	expression tag	UNP Q8JJ25
B	115	HIS	-	expression tag	UNP Q8JJ25
B	116	HIS	-	expression tag	UNP Q8JJ25
B	117	HIS	-	expression tag	UNP Q8JJ25

- Molecule 3 is ADENOSINE-5'-DIPHOSPHATE (three-letter code: ADP) (formula:  $C_{10}H_{15}N_5O_{10}P_2$ ).



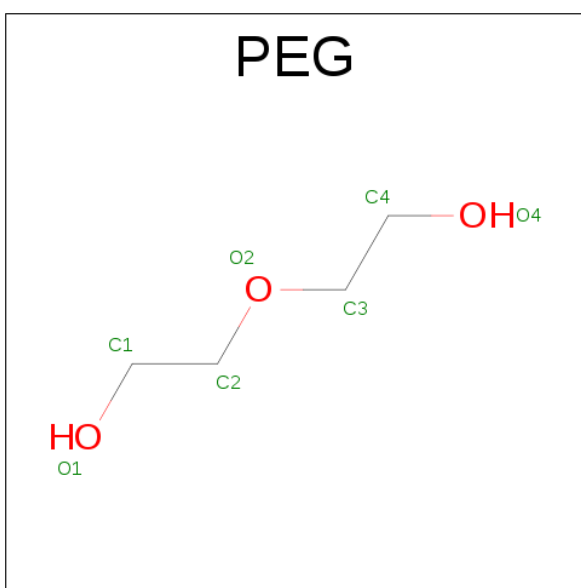
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	P	
			27	10	5	10	2	

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula:  $C_2H_6O_2$ ).



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

- Molecule 5 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula:  $C_4H_{10}O_3$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	C	O	0	0
			7	4	3		
5	A	1	Total	C	O	0	0
			7	4	3		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	139	Total 139	O 139	0	0
6	B	58	Total 58	O 58	0	0



- Molecule 1: Aurora kinase A





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	91.23Å 111.12Å 41.62Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	55.56 – 1.68 55.56 – 1.68	Depositor EDS
% Data completeness (in resolution range)	99.7 (55.56-1.68) 99.7 (55.56-1.68)	Depositor EDS
$R_{merge}$	0.04	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.59 (at 1.68Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.4_1496)	Depositor
R, $R_{free}$	0.199 , 0.231 0.200 , 0.233	Depositor DCC
$R_{free}$ test set	2449 reflections (5.00%)	DCC
Wilson B-factor (Å <sup>2</sup> )	29.0	Xtriage
Anisotropy	0.379	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.38 , 59.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	3041	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	36.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.32% 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, 0.333 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: PEG, EDO, ADP

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	0.39	0/2076	0.52	0/2819
2	B	0.35	0/778	0.53	0/1058
All	All	0.38	0/2854	0.52	0/3877

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	2024	0	1962	4	0
2	B	771	0	734	2	0
3	A	27	0	12	0	0
4	A	8	0	12	0	0
5	A	14	0	20	3	0
6	A	139	0	0	3	0
6	B	58	0	0	1	0
All	All	3041	0	2740	9	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 2.

All (9) 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)
5:A:504:PEG:H11	6:A:604:HOH:O	1.96	0.64
1:A:253[B]:ILE:HG13	1:A:278:SER:HB3	1.91	0.52
2:B:39:ARG:HD3	2:B:82:TYR:CZ	2.45	0.51
1:A:144:PHE:HD1	1:A:168:GLN:HB3	1.75	0.50
2:B:9:ARG:HD2	6:B:246:HOH:O	2.15	0.47
5:A:504:PEG:C2	6:A:604:HOH:O	2.63	0.47
1:A:307:ASP:O	1:A:310:VAL:HG22	2.17	0.45
5:A:504:PEG:C1	6:A:604:HOH:O	2.63	0.42
1:A:162:LYS:HE2	1:A:164:LEU:HD21	2.01	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	251/285 (88%)	244 (97%)	7 (3%)	0	100	100
2	B	103/117 (88%)	102 (99%)	1 (1%)	0	100	100
All	All	354/402 (88%)	346 (98%)	8 (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	208/248 (84%)	208 (100%)	0	100	100
2	B	83/97 (86%)	83 (100%)	0	100	100
All	All	291/345 (84%)	291 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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 ⓘ

5 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	$\# Z  > 2$
3	ADP	A	501	-	25,29,29	0.93	1 (4%)	24,45,45	1.82	2 (8%)
4	EDO	A	502	-	3,3,3	0.41	0	2,2,2	0.43	0
4	EDO	A	503	-	3,3,3	0.48	0	2,2,2	0.38	0
5	PEG	A	504	-	6,6,6	0.59	0	5,5,5	0.28	0
5	PEG	A	505	-	6,6,6	0.57	0	5,5,5	0.35	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	ADP	A	501	-	-	0/12/32/32	0/3/3/3
4	EDO	A	502	-	-	0/1/1/1	0/0/0/0
4	EDO	A	503	-	-	0/1/1/1	0/0/0/0
5	PEG	A	504	-	-	0/4/4/4	0/0/0/0
5	PEG	A	505	-	-	0/4/4/4	0/0/0/0

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	501	ADP	C5-C4	2.90	1.47	1.40

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	501	ADP	N3-C2-N1	-6.90	122.84	128.86
3	A	501	ADP	C2-N1-C6	2.23	122.67	118.77

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	504	PEG	3	0

## 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, 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	253/285 (88%)	-0.00	8 (3%)	48	52	23, 32, 53, 60	0
2	B	105/117 (89%)	0.15	5 (4%)	31	34	25, 36, 56, 70	0
All	All	358/402 (89%)	0.04	13 (3%)	43	47	23, 33, 55, 70	0

All (13) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	105	GLY	4.2
1	A	126	ARG	4.1
1	A	390	PRO	4.1
1	A	334	TYR	3.6
1	A	391	SER	3.2
1	A	292	THR	2.7
2	B	45	THR	2.5
2	B	15	THR	2.5
1	A	303	GLY	2.4
2	B	103	VAL	2.1
1	A	127	GLN	2.1
2	B	42	LEU	2.1
1	A	305	MET	2.0

### 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](#)

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> 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	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
5	PEG	A	505	7/7	0.83	0.13	3.97	46,53,55,63	0
4	EDO	A	502	4/4	0.96	0.12	0.04	34,35,46,47	0
3	ADP	A	501	27/27	0.92	0.09	0.03	27,42,63,64	0
4	EDO	A	503	4/4	0.81	0.09	-0.51	49,54,57,57	0
5	PEG	A	504	7/7	0.91	0.09	-0.68	37,42,49,49	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.