



# Full wwPDB X-ray Structure Validation Report ⓘ

May 24, 2020 – 01:27 am BST

PDB ID : 4YOM  
Title : Structure of SAD kinase  
Authors : Wu, J.X.; Wang, J.; Chen, L.; Wang, Z.X.; Wu, J.W.  
Deposited on : 2015-03-12  
Resolution : 2.49 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

<https://www.wwpdb.org/validation/2017/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 : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.11  
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) : 2.11

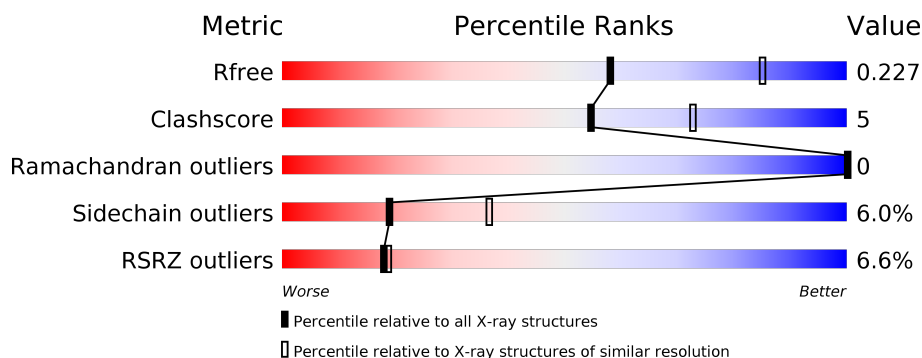
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.49 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))
$R_{free}$	130704	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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  $\leq 5\%$ . 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	B	350	<div> <div>5%</div> <div> <div></div> <div>77%</div> <div>13%</div> <div>8%</div> </div> </div>
2	A	144	<div> <div>8%</div> <div> <div></div> <div>59%</div> <div>13%</div> <div>28%</div> </div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 3515 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 Serine/threonine-protein kinase BRSK2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	B	322	2601	1661	460	465	15	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	343	LEU	-	expression tag	UNP Q69Z98
B	344	GLU	-	expression tag	UNP Q69Z98
B	345	HIS	-	expression tag	UNP Q69Z98
B	346	HIS	-	expression tag	UNP Q69Z98
B	347	HIS	-	expression tag	UNP Q69Z98
B	348	HIS	-	expression tag	UNP Q69Z98
B	349	HIS	-	expression tag	UNP Q69Z98
B	350	HIS	-	expression tag	UNP Q69Z98

- Molecule 2 is a protein called Serine/threonine-protein kinase BRSK2.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
2	A	104	839	546	140	153	0	0	0

There are 9 discrepancies between the modelled and reference sequences:

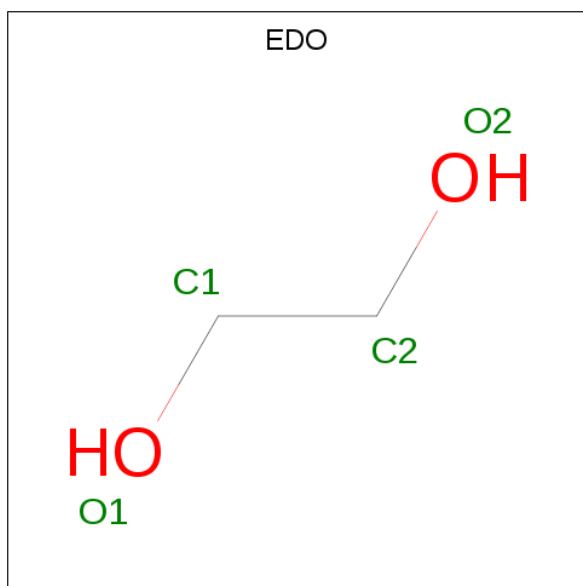
Chain	Residue	Modelled	Actual	Comment	Reference
A	518	MET	-	expression tag	UNP Q69Z98
A	654	LEU	-	expression tag	UNP Q69Z98
A	655	GLU	-	expression tag	UNP Q69Z98
A	656	HIS	-	expression tag	UNP Q69Z98
A	657	HIS	-	expression tag	UNP Q69Z98
A	658	HIS	-	expression tag	UNP Q69Z98
A	659	HIS	-	expression tag	UNP Q69Z98
A	660	HIS	-	expression tag	UNP Q69Z98

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	661	HIS	-	expression tag	UNP Q69Z98

- Molecule 3 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula:  $C_2H_6O_2$ ).



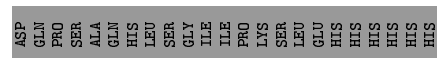
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total C O 4 2 2	0	0
3	B	1	Total C O 4 2 2	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	59	Total O 59 59	0	0
4	A	8	Total O 8 8	0	0



- Molecule 1: Serine/threonine-protein kinase BRSK2



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	98.11Å 87.27Å 80.30Å 90.00° 92.29° 90.00°	Depositor
Resolution (Å)	34.58 – 2.49 34.58 – 2.49	Depositor EDS
% Data completeness (in resolution range)	99.7 (34.58-2.49) 99.4 (34.58-2.49)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.42 (at 2.48Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7.3_928)	Depositor
R, $R_{free}$	0.197 , 0.236 0.192 , 0.227	Depositor DCC
$R_{free}$ test set	1217 reflections (5.11%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	56.8	Xtriage
Anisotropy	0.629	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 41.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.020 for -h,-k,l	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	3515	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.95% 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

### 5.1 Standard geometry

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

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	B	0.45	0/2651	0.57	0/3568
2	A	0.38	0/855	0.49	0/1150
All	All	0.44	0/3506	0.55	0/4718

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

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	B	2601	0	2661	29	0
2	A	839	0	862	9	0
3	B	8	0	12	0	0
4	A	8	0	0	0	0
4	B	59	0	0	2	0
All	All	3515	0	3535	36	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 5.

All (36) 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:53:ARG:HG3	1:B:53:ARG:HH11	1.41	0.83
1:B:115:THR:HG22	1:B:118:GLU:H	1.52	0.74
1:B:305:ASP:OD2	2:A:621:ARG:NH2	2.25	0.68
1:B:55:LYS:NZ	1:B:172:LEU:O	2.35	0.59
1:B:53:ARG:HG3	1:B:53:ARG:NH1	2.12	0.57
1:B:51:VAL:HG22	1:B:168:VAL:HG13	1.87	0.56
2:A:544:LEU:HG	2:A:548:LYS:HD2	1.88	0.56
1:B:189:ARG:NH2	4:B:541:HOH:O	2.42	0.52
1:B:194:ASP:HB3	1:B:197:LYS:HD2	1.92	0.51
2:A:566:ILE:HG21	2:A:572:ARG:HG3	1.93	0.51
1:B:113:ARG:HG3	1:B:211:LEU:O	2.11	0.50
2:A:588:VAL:HG21	2:A:622:ARG:HB3	1.95	0.49
1:B:307:MET:HE2	1:B:335:TYR:HD1	1.76	0.49
1:B:308:HIS:CE1	1:B:316:ARG:HG3	2.46	0.49
1:B:304:LEU:HD21	1:B:320:LEU:HD21	1.96	0.48
2:A:525:ASN:HA	2:A:624:LYS:HE3	1.96	0.48
1:B:303:VAL:HG12	1:B:307:MET:HE2	1.95	0.47
1:B:240:ILE:O	1:B:245:GLN:NE2	2.37	0.46
1:B:136:HIS:HD2	1:B:333:MET:HE1	1.81	0.46
1:B:53:ARG:HH11	1:B:53:ARG:CG	2.20	0.46
1:B:192:LYS:HA	1:B:192:LYS:HD3	1.81	0.45
2:A:548:LYS:HE2	2:A:568:GLN:O	2.17	0.45
1:B:197:LYS:HE3	4:B:556:HOH:O	2.16	0.45
1:B:302:ASP:HA	2:A:621:ARG:HH22	1.82	0.44
1:B:62:MET:HG2	1:B:312:CYS:HA	1.98	0.44
1:B:290:ILE:HD11	1:B:333:MET:HG2	2.00	0.43
1:B:307:MET:CE	1:B:335:TYR:HD1	2.32	0.42
2:A:624:LYS:O	2:A:628:GLU:HG3	2.19	0.42
1:B:315:ASP:HB3	1:B:318:LYS:HB3	2.02	0.42
1:B:170:ASP:OD1	1:B:171:SER:N	2.52	0.42
1:B:172:LEU:HD22	1:B:172:LEU:HA	1.85	0.42
1:B:304:LEU:HD12	1:B:304:LEU:HA	1.84	0.41
1:B:288:VAL:HB	1:B:337:LEU:HD23	2.03	0.41
1:B:294:PRO:HG2	1:B:298:ASP:OD2	2.21	0.41
2:A:558:ILE:HA	2:A:559:PRO:HD2	1.95	0.41
1:B:81:LYS:HD3	1:B:83:HIS:CE1	2.57	0.40

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	B	318/350 (91%)	310 (98%)	8 (2%)	0	100	100
2	A	98/144 (68%)	95 (97%)	3 (3%)	0	100	100
All	All	416/494 (84%)	405 (97%)	11 (3%)	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	B	289/312 (93%)	273 (94%)	16 (6%)	21	41
2	A	96/128 (75%)	89 (93%)	7 (7%)	14	27
All	All	385/440 (88%)	362 (94%)	23 (6%)	19	37

All (23) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	30	GLN
1	B	53	ARG
1	B	72	LYS
1	B	81	LYS
1	B	168	VAL
1	B	172	LEU
1	B	173	LEU
1	B	224	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	231	ARG
1	B	263	GLU
1	B	275	LYS
1	B	276	ASN
1	B	296	LEU
1	B	297	GLU
1	B	343	LEU
1	B	344	GLU
2	A	560	SER
2	A	567	SER
2	A	598	GLU
2	A	606	ASN
2	A	621	ARG
2	A	629	THR
2	A	636	SER

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

Mol	Chain	Res	Type
1	B	83	HIS
2	A	606	ASN

### 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](#)

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	$\# Z  > 2$
3	EDO	B	402	-	3,3,3	0.53	0	2,2,2	0.27	0
3	EDO	B	401	-	3,3,3	0.45	0	2,2,2	0.58	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	EDO	B	402	-	-	1/1/1/1	-
3	EDO	B	401	-	-	1/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	B	402	EDO	O1-C1-C2-O2
3	B	401	EDO	O1-C1-C2-O2

There are no ring outliers.

No monomer is involved in short contacts.

## 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	B	322/350 (92%)	0.21	16 (4%) 28 30	20, 44, 109, 174	0
2	A	104/144 (72%)	0.32	12 (11%) 4 4	34, 59, 99, 132	0
All	All	426/494 (86%)	0.24	28 (6%) 18 19	20, 47, 106, 174	0

All (28) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	A	585	GLN	5.2
1	B	291	ARG	4.3
2	A	576	LYS	4.0
2	A	586	LYS	3.9
1	B	31	THR	3.8
1	B	344	GLU	3.4
2	A	599	GLY	3.4
1	B	13	HIS	3.4
2	A	622	ARG	3.1
1	B	326	GLU	3.1
2	A	540	LYS	2.8
1	B	89	LYS	2.8
1	B	172	LEU	2.8
2	A	541	ASP	2.7
1	B	289	GLN	2.7
1	B	110	LYS	2.7
1	B	343	LEU	2.6
1	B	290	ILE	2.4
1	B	109	VAL	2.4
1	B	230	LYS	2.3
1	B	276	ASN	2.3
2	A	621	ARG	2.3
2	A	575	TYR	2.3
1	B	178	GLY	2.2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
2	A	596	TYR	2.2
2	A	587	PRO	2.1
2	A	598	GLU	2.1
1	B	297	GLU	2.1

## 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. 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	B-factors( $\text{\AA}^2$ )	Q<0.9
3	EDO	B	402	4/4	0.88	0.25	65,65,65,65	0
3	EDO	B	401	4/4	0.92	0.19	61,61,61,61	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.