Full wwPDB X-ray Structure Validation Report (i)

Jun 18, 2024 – 06:39 PM EDT

PDB ID	:	4I93
Title	:	Structure of the BSK8 kinase domain (SeMet labeled)
Authors	:	Gruetter, C.; Rauh, D.
Deposited on	:	2012-12-04
Resolution	:	1.50 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

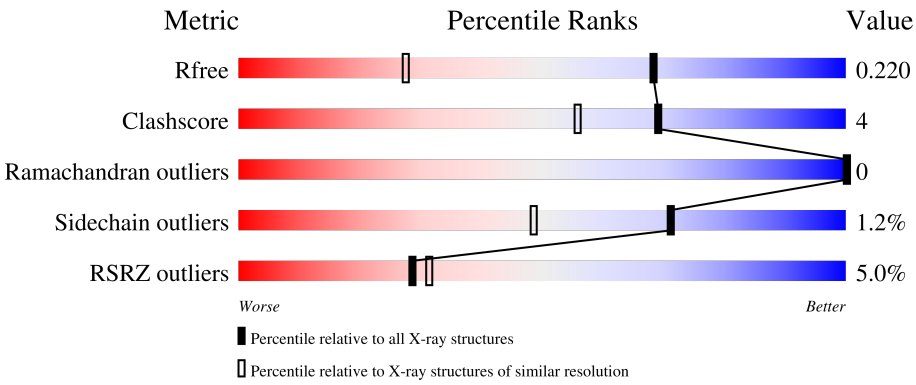
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.50 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	2936 (1.50-1.50)
Clashscore	141614	3144 (1.50-1.50)
Ramachandran outliers	138981	3066 (1.50-1.50)
Sidechain outliers	138945	3064 (1.50-1.50)
RSRZ outliers	127900	2884 (1.50-1.50)

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 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	300	3% 86% 10% • •
1	B	300	6% 88% 6% 6%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 5058 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 Probable serine/threonine-protein kinase At5g41260.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	289	2363	1492	424	432	5	10	0	12	0
1	B	283	2281	1440	406	421	5	9	0	7	0

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	29	MSE	-	EXPRESSION TAG	UNP Q9FHD7
A	30	ALA	-	EXPRESSION TAG	UNP Q9FHD7
A	31	HIS	-	EXPRESSION TAG	UNP Q9FHD7
A	32	HIS	-	EXPRESSION TAG	UNP Q9FHD7
A	33	HIS	-	EXPRESSION TAG	UNP Q9FHD7
A	34	HIS	-	EXPRESSION TAG	UNP Q9FHD7
A	35	HIS	-	EXPRESSION TAG	UNP Q9FHD7
A	36	HIS	-	EXPRESSION TAG	UNP Q9FHD7
A	37	SER	-	EXPRESSION TAG	UNP Q9FHD7
A	38	SER	-	EXPRESSION TAG	UNP Q9FHD7
A	39	GLY	-	EXPRESSION TAG	UNP Q9FHD7
B	29	MSE	-	EXPRESSION TAG	UNP Q9FHD7
B	30	ALA	-	EXPRESSION TAG	UNP Q9FHD7
B	31	HIS	-	EXPRESSION TAG	UNP Q9FHD7
B	32	HIS	-	EXPRESSION TAG	UNP Q9FHD7
B	33	HIS	-	EXPRESSION TAG	UNP Q9FHD7
B	34	HIS	-	EXPRESSION TAG	UNP Q9FHD7
B	35	HIS	-	EXPRESSION TAG	UNP Q9FHD7
B	36	HIS	-	EXPRESSION TAG	UNP Q9FHD7
B	37	SER	-	EXPRESSION TAG	UNP Q9FHD7
B	38	SER	-	EXPRESSION TAG	UNP Q9FHD7
B	39	GLY	-	EXPRESSION TAG	UNP Q9FHD7

- Molecule 2 is water.

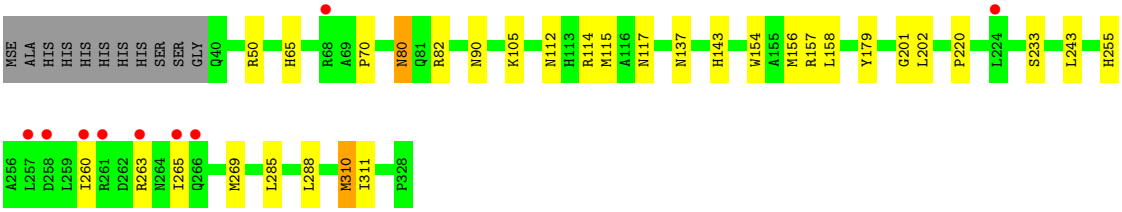
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	248	248	248	0	0
			Total	O		
2	B	166	166	166	0	0

3 Residue-property plots (i)

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.

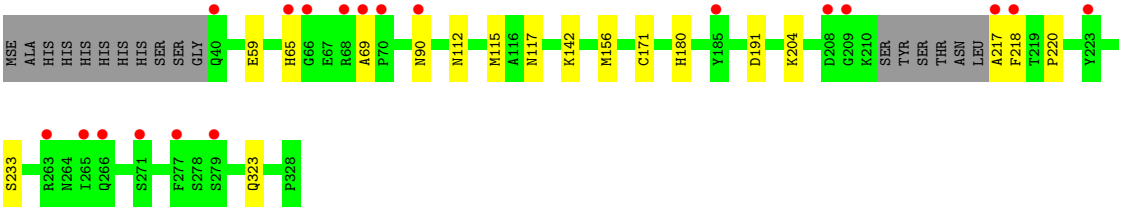
• Molecule 1: Probable serine/threonine-protein kinase At5g41260

Chain A:



• Molecule 1: Probable serine/threonine-protein kinase At5g41260

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	50.90Å 77.55Å 68.15Å 90.00° 105.67° 90.00°	Depositor
Resolution (Å)	33.39 – 1.50 33.38 – 1.50	Depositor EDS
% Data completeness (in resolution range)	98.1 (33.39-1.50) 98.1 (33.38-1.50)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	5.63 (at 1.50Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.180 , 0.208 0.190 , 0.220	Depositor DCC
R_{free} test set	1000 reflections (1.25%)	wwPDB-VP
Wilson B-factor (Å^2)	13.9	Xtriage
Anisotropy	0.251	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.41 , 46.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	5058	wwPDB-VP
Average B, all atoms (Å^2)	19.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 28.82 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.7227e-03.*

[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	0.58	0/2439	0.80	3/3277 (0.1%)
1	B	0.53	0/2340	0.73	1/3148 (0.0%)
All	All	0.56	0/4779	0.77	4/6425 (0.1%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	156	MSE	CG-SE-CE	-8.82	79.50	98.90
1	B	156	MSE	CG-SE-CE	-5.94	85.83	98.90
1	A	157	ARG	NE-CZ-NH2	-5.37	117.61	120.30
1	A	50	ARG	NE-CZ-NH1	5.04	122.82	120.30

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	2363	0	2368	21	0
1	B	2281	0	2248	14	0
2	A	248	0	0	5	0
2	B	166	0	0	1	0
All	All	5058	0	4616	34	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 4.

All (34) 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:310:MSE:HA	1:A:310:MSE:HE2	1.41	1.00
1:A:310:MSE:SE	2:A:505:HOH:O	2.39	0.91
1:A:220:PRO:HG2	1:A:233[A]:SER:HB2	1.61	0.83
1:B:115[B]:MSE:HE1	1:B:171:CYS:SG	2.19	0.83
1:A:310:MSE:HA	1:A:310:MSE:CE	2.19	0.71
1:B:115[B]:MSE:CE	1:B:171:CYS:SG	2.83	0.66
1:A:65:HIS:ND1	2:A:623:HOH:O	2.28	0.61
1:A:105:LYS:NZ	2:A:646:HOH:O	2.33	0.59
1:A:112:ASN:H	1:A:117:ASN:HD21	1.51	0.58
1:A:288:LEU:HD22	1:A:310:MSE:HE3	1.86	0.58
1:B:180:HIS:CE1	1:B:204:LYS:HG2	2.41	0.55
1:B:220:PRO:HG2	1:B:233[A]:SER:HB2	1.91	0.52
1:B:323[A]:GLN:NE2	2:B:464:HOH:O	2.43	0.52
1:A:265:ILE:HA	1:A:269:MSE:HG2	1.92	0.51
1:B:65:HIS:HB2	1:B:69:ALA:HB2	1.93	0.50
1:A:112:ASN:ND2	1:A:114:ARG:H	2.10	0.50
1:A:137:ASN:O	1:A:143:HIS:HE1	1.94	0.50
1:A:115[B]:MSE:HE1	1:A:202:LEU:HD11	1.94	0.49
1:B:218:PHE:CD1	1:B:218:PHE:N	2.80	0.48
1:B:65:HIS:CB	1:B:69:ALA:HB2	2.45	0.46
1:A:311:ILE:HG21	1:B:217:ALA:HA	1.97	0.46
1:B:217:ALA:C	1:B:218:PHE:CD1	2.89	0.46
1:A:179:TYR:CZ	1:A:201:GLY:HA3	2.52	0.45
1:A:285:LEU:CD1	1:A:310:MSE:HE1	2.47	0.44
1:B:112:ASN:H	1:B:117:ASN:HD21	1.64	0.44
1:A:70:PRO:HA	1:A:90:ASN:HD21	1.83	0.44
1:A:263:ARG:NH2	2:A:575:HOH:O	2.51	0.43
1:B:59:GLU:CD	1:B:59:GLU:H	2.22	0.43
1:B:217:ALA:C	1:B:218:PHE:HD1	2.22	0.42
1:A:143:HIS:HD2	2:A:609:HOH:O	2.01	0.42
1:A:80:ASN:ND2	1:A:82:ARG:H	2.17	0.42
1:A:154:TRP:NE1	1:A:158[A]:LEU:HD11	2.35	0.41
1:B:218:PHE:N	1:B:218:PHE:HD1	2.18	0.41
1:A:243:LEU:HD21	1:A:260:ILE:HD11	2.03	0.40

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	299/300 (100%)	295 (99%)	4 (1%)	0	100	100
1	B	286/300 (95%)	284 (99%)	2 (1%)	0	100	100
All	All	585/600 (98%)	579 (99%)	6 (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	260/252 (103%)	257 (99%)	3 (1%)	71	48
1	B	246/252 (98%)	243 (99%)	3 (1%)	71	48
All	All	506/504 (100%)	500 (99%)	6 (1%)	71	48

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	80	ASN
1	A	255	HIS
1	A	310	MSE
1	B	90	ASN
1	B	142	LYS
1	B	191	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (8) such

sidechains are listed below:

Mol	Chain	Res	Type
1	A	80	ASN
1	A	90	ASN
1	A	112	ASN
1	A	117	ASN
1	A	143	HIS
1	B	60	ASN
1	B	117	ASN
1	B	180	HIS

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 monosaccharides 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, 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	281/300 (93%)	0.19	9 (3%)	47	52	7, 14, 29, 49	0
1	B	275/300 (91%)	0.46	19 (6%)	16	17	9, 17, 42, 76	0
All	All	556/600 (92%)	0.32	28 (5%)	28	31	7, 16, 36, 76	0

All (28) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	217	ALA	5.7
1	B	65	HIS	5.6
1	B	68	ARG	5.5
1	B	208	ASP	4.7
1	B	209	GLY	4.5
1	B	66	GLY	4.2
1	A	265	ILE	4.0
1	B	69	ALA	3.6
1	B	40	GLN	3.5
1	A	261	ARG	3.3
1	B	70	PRO	3.3
1	B	90	ASN	3.1
1	B	265	ILE	3.0
1	A	266	GLN	2.9
1	A	263	ARG	2.9
1	A	260	ILE	2.9
1	B	263	ARG	2.6
1	A	257	LEU	2.6
1	A	224	LEU	2.6
1	A	68	ARG	2.4
1	A	258	ASP	2.4
1	B	218	PHE	2.4
1	B	277	PHE	2.3
1	B	223	TYR	2.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	271	SER	2.2
1	B	279	SER	2.1
1	B	266	GLN	2.1
1	B	185[A]	TYR	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 monosaccharides 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.