



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 25, 2020 – 03:39 pm BST

PDB ID : 5L2Q  
Title : Serine/threonine-protein kinase 40 (STK40) kinase homology domain  
Authors : Durzynska, I.; Uljon, S.; Blacklow, S.C.  
Deposited on : 2016-08-02  
Resolution : 2.53 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

<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
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

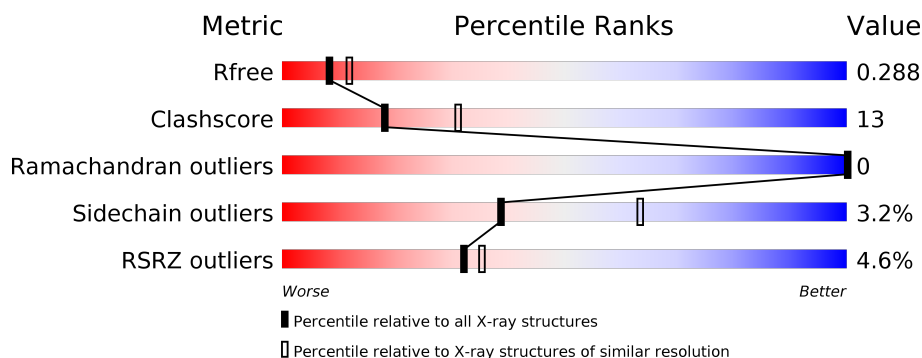
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.53 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	5743 (2.54-2.50)
Clashscore	141614	6463 (2.54-2.50)
Ramachandran outliers	138981	6335 (2.54-2.50)
Sidechain outliers	138945	6337 (2.54-2.50)
RSRZ outliers	127900	5630 (2.54-2.50)

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	A	320	<div> <div> <div>0%</div> <div>65%</div> <div>25%</div> <div>8%</div> </div> </div>
1	B	320	<div> <div>2%</div> <div>66%</div> <div>22%</div> <div>12%</div> </div>
1	C	320	<div> <div>7%</div> <div>67%</div> <div>21%</div> <div>12%</div> </div>
1	D	320	<div> <div>7%</div> <div>67%</div> <div>24%</div> <div>8%</div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 8350 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 40.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	282	Total 2023	C 1298	N 352	O 362	S 11	5	1	0
1	B	282	Total 2072	C 1320	N 361	O 381	S 10	0	0	0
1	A	293	Total 2166	C 1394	N 376	O 386	S 10	0	0	0
1	D	294	Total 2023	C 1273	N 358	O 381	S 11	0	0	0

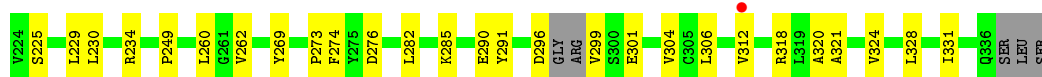
There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	20	GLY	-	expression tag	UNP Q8N2I9
A	21	SER	-	expression tag	UNP Q8N2I9
B	20	GLY	-	expression tag	UNP Q8N2I9
B	21	SER	-	expression tag	UNP Q8N2I9
C	20	GLY	-	expression tag	UNP Q8N2I9
C	21	SER	-	expression tag	UNP Q8N2I9
D	20	GLY	-	expression tag	UNP Q8N2I9
D	21	SER	-	expression tag	UNP Q8N2I9

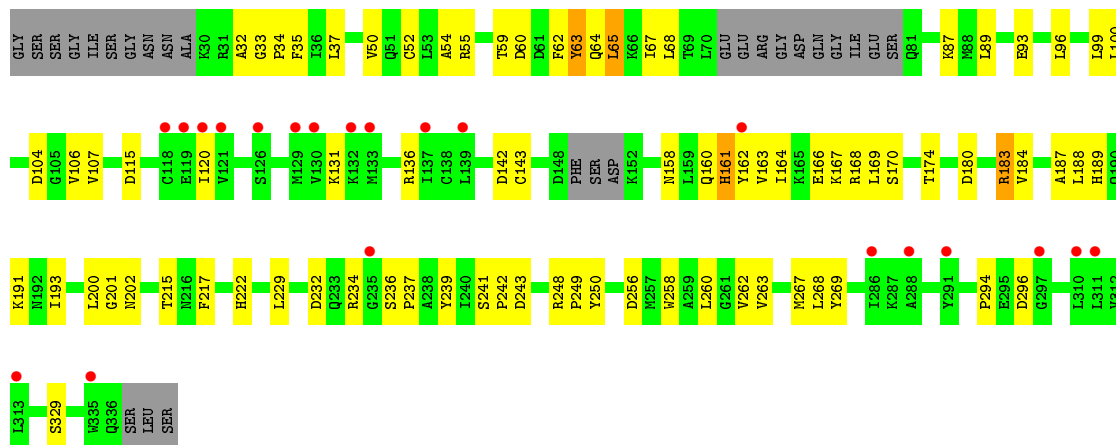
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	C	16	Total 16	O 16	0	0
2	B	17	Total 17	O 17	0	0
2	A	20	Total 20	O 20	0	0
2	D	13	Total 13	O 13	0	0





- Molecule 1: Serine/threonine-protein kinase 40



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	55.55Å 237.65Å 55.61Å 90.00° 106.32° 90.00°	Depositor
Resolution (Å)	59.41 – 2.53 59.41 – 2.53	Depositor EDS
% Data completeness (in resolution range)	98.6 (59.41-2.53) 98.6 (59.41-2.53)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	47.95 (at 2.55Å)	Xtriage
Refinement program	PHENIX 1.11.1_2575	Depositor
R, $R_{free}$	0.258 , 0.280 0.267 , 0.288	Depositor DCC
$R_{free}$ test set	2383 reflections (5.27%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	62.4	Xtriage
Anisotropy	0.350	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.24 , 55.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.45$ , $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	0.043 for l,-k,h	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	8350	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	83.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.49% 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 ⓘ

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	0.56	0/2208	0.54	0/3005
1	B	0.57	1/2114 (0.0%)	0.52	0/2882
1	C	0.35	0/2058	0.55	0/2810
1	D	0.44	0/2060	0.60	0/2828
All	All	0.49	1/8440 (0.0%)	0.55	0/11525

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	162	TYR	CE1-CZ	-5.10	1.31	1.38

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	A	2166	0	2083	55	0
1	B	2072	0	1932	55	0
1	C	2023	0	1873	45	0
1	D	2023	0	1718	59	0
2	A	20	0	0	2	0
2	B	17	0	0	2	0
2	C	16	0	0	0	0
2	D	13	0	0	1	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	8350	0	7606	213	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 13.

The worst 5 of 213 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:D:241:SER:CB	1:D:242:PRO:CD	2.19	1.16
1:B:303:THR:HG22	1:B:331:ILE:HD13	1.38	1.06
1:B:94:TYR:CE1	1:B:109:HIS:HE1	1.78	1.01
1:D:241:SER:CB	1:D:242:PRO:HD2	1.91	1.00
1:D:241:SER:CB	1:D:242:PRO:HD3	1.94	0.97

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	285/320 (89%)	271 (95%)	14 (5%)	0	100	100
1	B	276/320 (86%)	264 (96%)	12 (4%)	0	100	100
1	C	275/320 (86%)	259 (94%)	16 (6%)	0	100	100
1	D	288/320 (90%)	260 (90%)	28 (10%)	0	100	100
All	All	1124/1280 (88%)	1054 (94%)	70 (6%)	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	215/283 (76%)	208 (97%)	7 (3%)	38	62
1	B	204/283 (72%)	201 (98%)	3 (2%)	65	83
1	C	188/283 (66%)	184 (98%)	4 (2%)	53	76
1	D	172/283 (61%)	161 (94%)	11 (6%)	17	31
All	All	779/1132 (69%)	754 (97%)	25 (3%)	39	63

5 of 25 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	225	SER
1	A	296	ASP
1	D	256	ASP
1	A	230	LEU
1	D	63	TYR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	190	GLN
1	A	161	HIS
1	A	216	ASN
1	B	147	HIS
1	A	189	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 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	293/320 (91%)	-0.00	3 (1%) 82 84	39, 68, 107, 145	0
1	B	282/320 (88%)	0.15	6 (2%) 63 67	39, 71, 125, 171	0
1	C	282/320 (88%)	0.44	23 (8%) 11 12	39, 86, 144, 189	0
1	D	294/320 (91%)	0.34	21 (7%) 16 17	64, 102, 137, 166	0
All	All	1151/1280 (89%)	0.23	53 (4%) 32 35	39, 83, 134, 189	0

The worst 5 of 53 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	149	PHE	4.8
1	C	112	LEU	4.7
1	C	137	ILE	4.5
1	B	118	CYS	4.3
1	C	306	LEU	4.2

### 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.