



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 16, 2022 – 12:06 PM JST

PDB ID : 7VXG  
Title : Non-receptor Protein Tyrosine Phosphatase SHP2 in Complex with Allosteric Inhibitor TK-453  
Authors : Li, T.H.; Guo, H.T.; Ji, X.Y.  
Deposited on : 2021-11-12  
Resolution : 2.10 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  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.27  
buster-report : 1.1.7 (2018)  
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) : 2.27

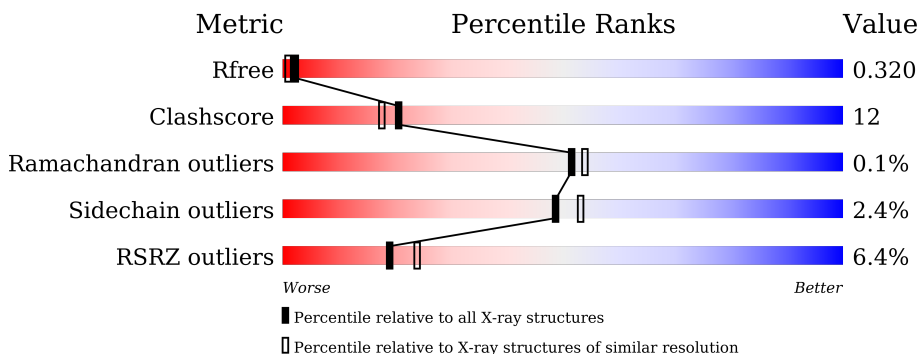
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.10 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	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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  $\geq 3$ , 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	526	<div> <div>5%</div> <div> <div>66%</div> <div>24%</div> <div>• 9%</div> </div> </div>
1	B	526	<div> <div>5%</div> <div> <div>73%</div> <div>17%</div> <div>• 9%</div> </div> </div>
1	C	526	<div> <div>6%</div> <div> <div>68%</div> <div>22%</div> <div>• 9%</div> </div> </div>
1	D	526	<div> <div>7%</div> <div> <div>69%</div> <div>23%</div> <div>• 7%</div> </div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 29981 atoms, of which 14395 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.

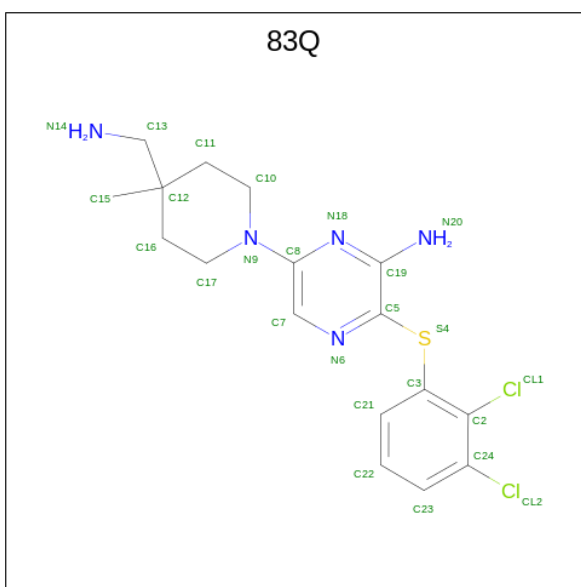
- Molecule 1 is a protein called Tyrosine-protein phosphatase non-receptor type 11.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	477	Total	C	H	N	O	S	0	0	0
			7338	2371	3581	653	715	18			
1	B	480	Total	C	H	N	O	S	0	0	0
			7409	2387	3624	666	715	17			
1	C	481	Total	C	H	N	O	S	0	0	0
			7213	2352	3478	649	718	16			
1	D	487	Total	C	H	N	O	S	0	0	0
			7457	2410	3628	674	728	17			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	SER	-	expression tag	UNP Q06124
B	0	SER	-	expression tag	UNP Q06124
C	0	SER	-	expression tag	UNP Q06124
D	0	SER	-	expression tag	UNP Q06124

- Molecule 2 is 6-[4-(aminomethyl)-4-methyl-piperidin-1-yl]-3-[2,3-bis(chloranyl)phenyl]sulfa nyl-pyrazin-2-amine (three-letter code: 83Q) (formula: C<sub>17</sub>H<sub>21</sub>Cl<sub>2</sub>N<sub>5</sub>S) (labeled as "Ligand of Interest" by depositor).



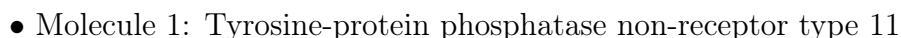
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
2	A	1	Total 46	C 17	Cl 2	H 21	N 5	S 1	0	0
2	B	1	Total 46	C 17	Cl 2	H 21	N 5	S 1	0	0
2	C	1	Total 46	C 17	Cl 2	H 21	N 5	S 1	0	0
2	D	1	Total 46	C 17	Cl 2	H 21	N 5	S 1	0	0

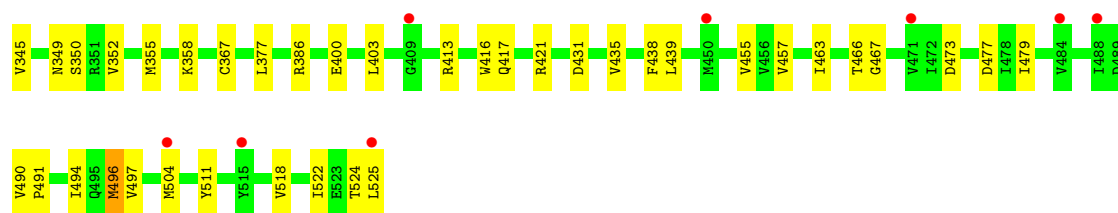
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	112	Total	O	0	0
			112	112		
3	B	87	Total	O	0	0
			87	87		
3	C	85	Total	O	0	0
			85	85		
3	D	96	Total	O	0	0
			96	96		

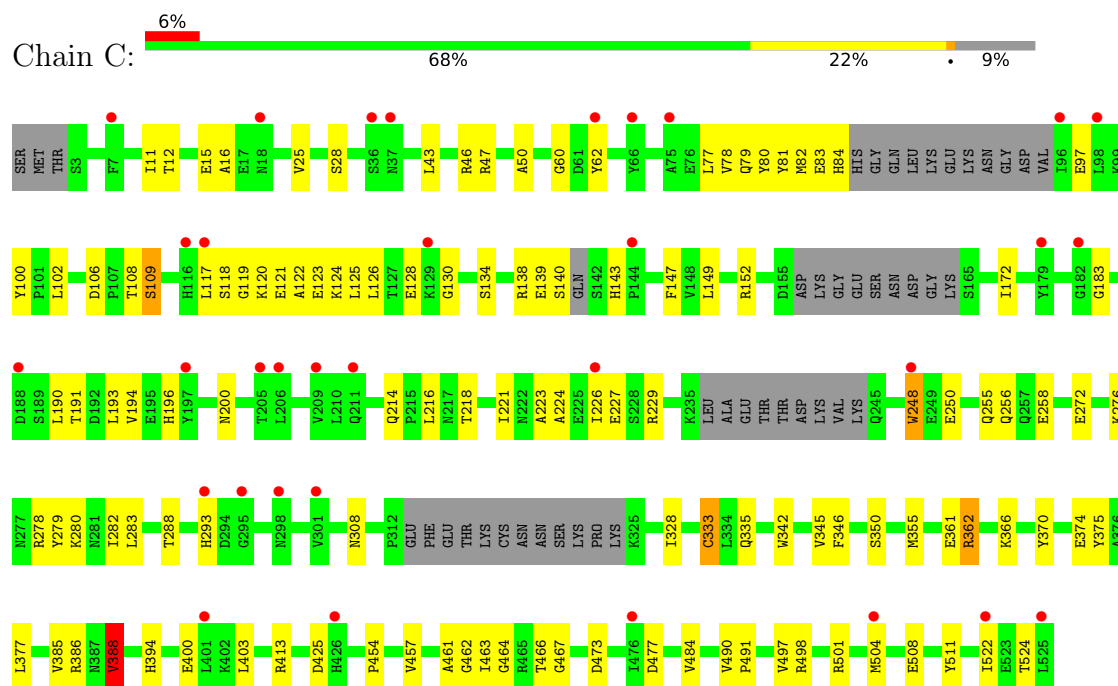


- Molecule 1: Tyrosine-protein phosphatase non-receptor type 11

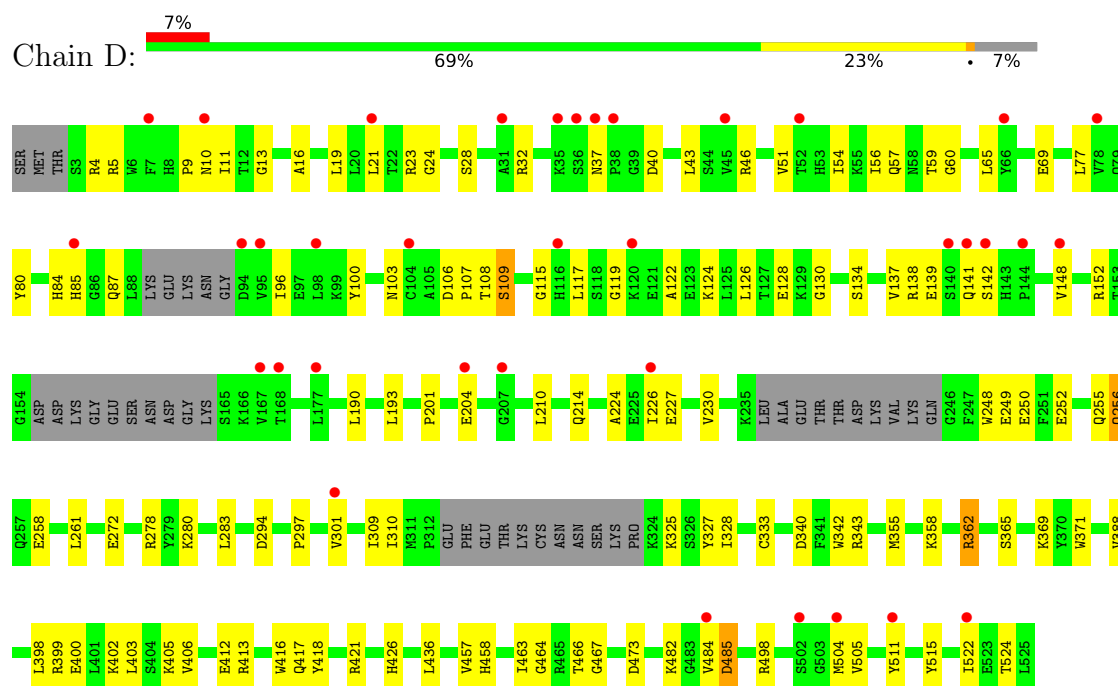




• Molecule 1: Tyrosine-protein phosphatase non-receptor type 11



• Molecule 1: Tyrosine-protein phosphatase non-receptor type 11



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	45.40Å 55.60Å 218.84Å 88.66° 102.01° 96.19°	Depositor
Resolution (Å)	38.14 – 2.10 38.14 – 2.10	Depositor EDS
% Data completeness (in resolution range)	83.7 (38.14-2.10) 83.6 (38.14-2.10)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.46 (at 2.10Å)	Xtriage
Refinement program	PHENIX (1.19.2_4158: ???)	Depositor
R, $R_{free}$	0.270 , 0.325 0.271 , 0.320	Depositor DCC
$R_{free}$ test set	1994 reflections (1.96%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	27.2	Xtriage
Anisotropy	0.528	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.41 , 36.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.430 for -h,-k,2*h+1	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	29981	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.54% 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: 83Q

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	0.67	1/3835 (0.0%)	0.76	1/5195 (0.0%)
1	B	0.67	1/3864 (0.0%)	0.79	2/5232 (0.0%)
1	C	0.68	2/3815 (0.1%)	0.78	0/5184
1	D	0.67	1/3911 (0.0%)	0.78	1/5302 (0.0%)
All	All	0.67	5/15425 (0.0%)	0.78	4/20913 (0.0%)

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	297	PRO	N-CD	5.42	1.55	1.47
1	A	327	TYR	CD2-CE2	5.33	1.47	1.39
1	C	333	CYS	CB-SG	-5.16	1.73	1.81
1	C	388	VAL	CB-CG2	-5.16	1.42	1.52
1	B	69	GLU	CD-OE1	-5.15	1.20	1.25

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	333	CYS	N-CA-CB	5.95	121.31	110.60
1	B	303	ASP	CB-CG-OD1	5.20	122.98	118.30
1	D	436	LEU	CB-CG-CD1	-5.04	102.43	111.00
1	A	383	MET	CG-SD-CE	5.01	108.22	100.20

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	3757	3581	3577	93	0
1	B	3785	3624	3620	76	0
1	C	3735	3478	3486	96	0
1	D	3829	3628	3636	98	0
2	A	25	21	0	2	0
2	B	25	21	0	1	0
2	C	25	21	0	1	0
2	D	25	21	0	1	0
3	A	112	0	0	3	0
3	B	87	0	0	4	0
3	C	85	0	0	1	0
3	D	96	0	0	2	0
All	All	15586	14395	14319	362	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 12.

The worst 5 of 362 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:4:ARG:NH1	1:D:258:GLU:OE2	1.92	1.01
1:B:4:ARG:NH1	1:B:258:GLU:OE2	1.96	0.98
1:B:249:GLU:OE2	2:B:601:83Q:N14	1.96	0.98
1:C:109:SER:OG	1:C:256:GLN:NE2	2.02	0.93
1:D:130:GLY:O	1:D:152:ARG:NH1	2.06	0.88

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	461/526 (88%)	445 (96%)	16 (4%)	0	100	100
1	B	466/526 (89%)	452 (97%)	13 (3%)	1 (0%)	47	49
1	C	469/526 (89%)	448 (96%)	21 (4%)	0	100	100
1	D	477/526 (91%)	458 (96%)	19 (4%)	0	100	100
All	All	1873/2104 (89%)	1803 (96%)	69 (4%)	1 (0%)	51	54

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	333	CYS

### 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	401/468 (86%)	388 (97%)	13 (3%)	39	41
1	B	403/468 (86%)	393 (98%)	10 (2%)	47	52
1	C	391/468 (84%)	382 (98%)	9 (2%)	50	55
1	D	406/468 (87%)	400 (98%)	6 (2%)	65	71
All	All	1601/1872 (86%)	1563 (98%)	38 (2%)	49	53

5 of 38 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	335	GLN
1	D	362	ARG
1	C	362	ARG
1	D	109	SER
1	D	485	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such

sidechains are listed below:

Mol	Chain	Res	Type
1	D	269	GLN
1	C	426	HIS
1	C	256	GLN
1	B	417	GLN
1	C	308	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 monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

4 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	$\# Z  > 2$
2	83Q	B	601	-	25,27,27	2.83	14 (56%)	29,39,39	2.01	8 (27%)
2	83Q	C	601	-	25,27,27	2.51	8 (32%)	29,39,39	1.75	8 (27%)
2	83Q	A	601	-	25,27,27	1.88	4 (16%)	29,39,39	1.92	7 (24%)
2	83Q	D	601	-	25,27,27	2.39	10 (40%)	29,39,39	1.66	7 (24%)

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
2	83Q	B	601	-	-	4/11/23/23	0/3/3/3
2	83Q	C	601	-	-	1/11/23/23	0/3/3/3
2	83Q	A	601	-	-	1/11/23/23	0/3/3/3
2	83Q	D	601	-	-	1/11/23/23	0/3/3/3

The worst 5 of 36 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	601	83Q	C10-N9	7.36	1.58	1.46
2	D	601	83Q	C2-CL1	-6.66	1.58	1.72
2	B	601	83Q	C2-CL1	-6.40	1.59	1.72
2	C	601	83Q	C24-CL2	-6.23	1.58	1.73
2	C	601	83Q	C10-N9	6.10	1.56	1.46

The worst 5 of 30 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	601	83Q	C24-C2-CL1	-5.16	113.38	120.02
2	B	601	83Q	C24-C2-CL1	-4.27	114.52	120.02
2	B	601	83Q	C3-S4-C5	4.16	114.88	102.91
2	D	601	83Q	C24-C2-CL1	-4.14	114.69	120.02
2	B	601	83Q	N18-C8-N9	4.07	121.11	116.55

There are no chirality outliers.

5 of 7 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	601	83Q	N6-C5-S4-C3
2	B	601	83Q	C15-C12-C13-N14
2	B	601	83Q	C11-C12-C13-N14
2	B	601	83Q	C16-C12-C13-N14
2	B	601	83Q	N6-C5-S4-C3

There are no ring outliers.

4 monomers are involved in 5 short contacts:

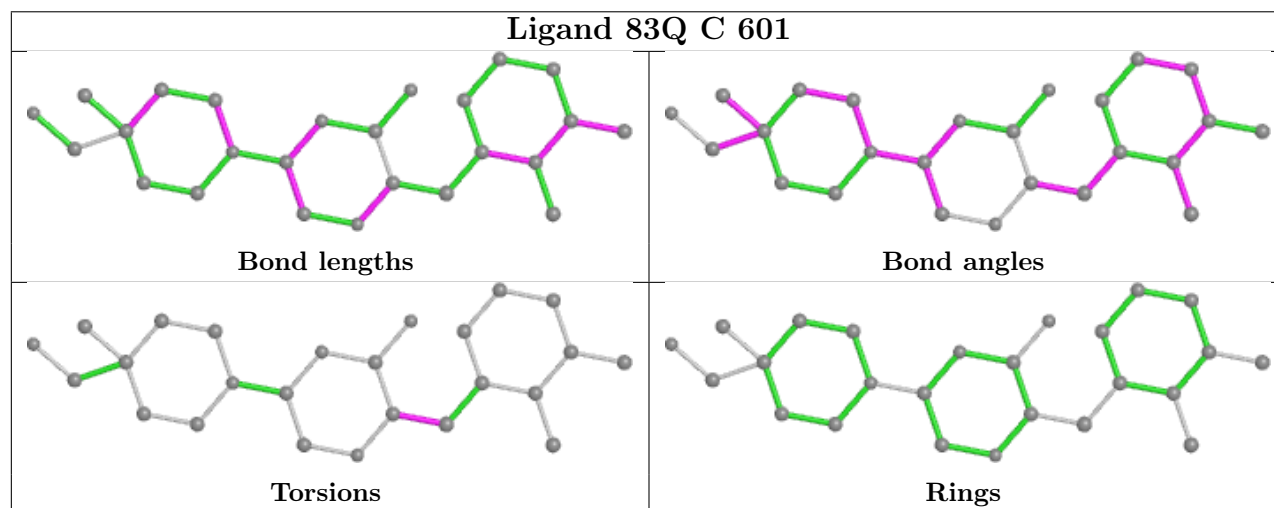
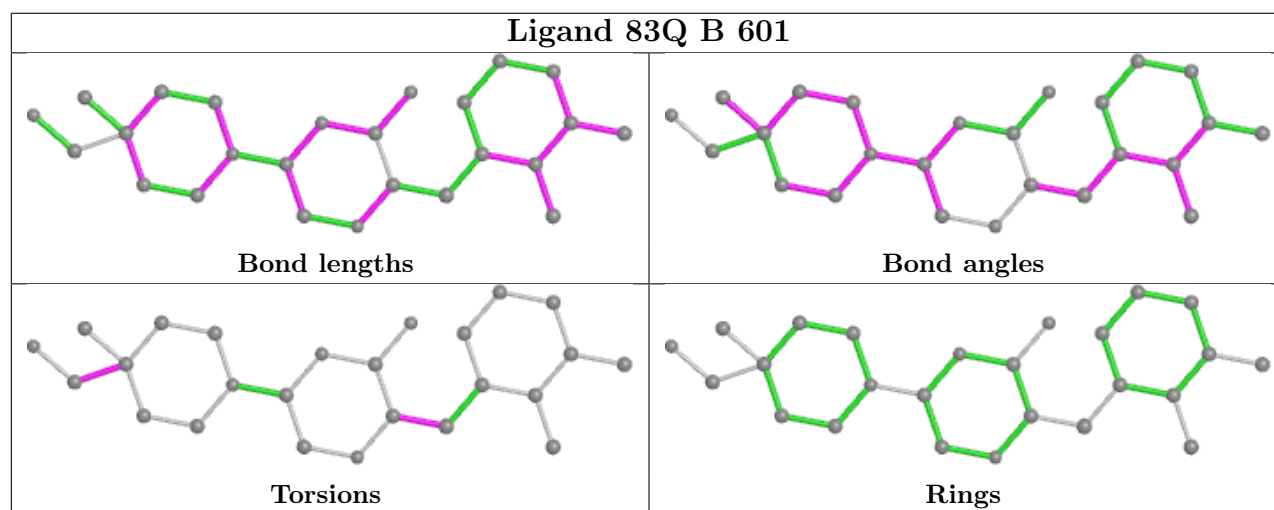
Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	601	83Q	1	0
2	C	601	83Q	1	0

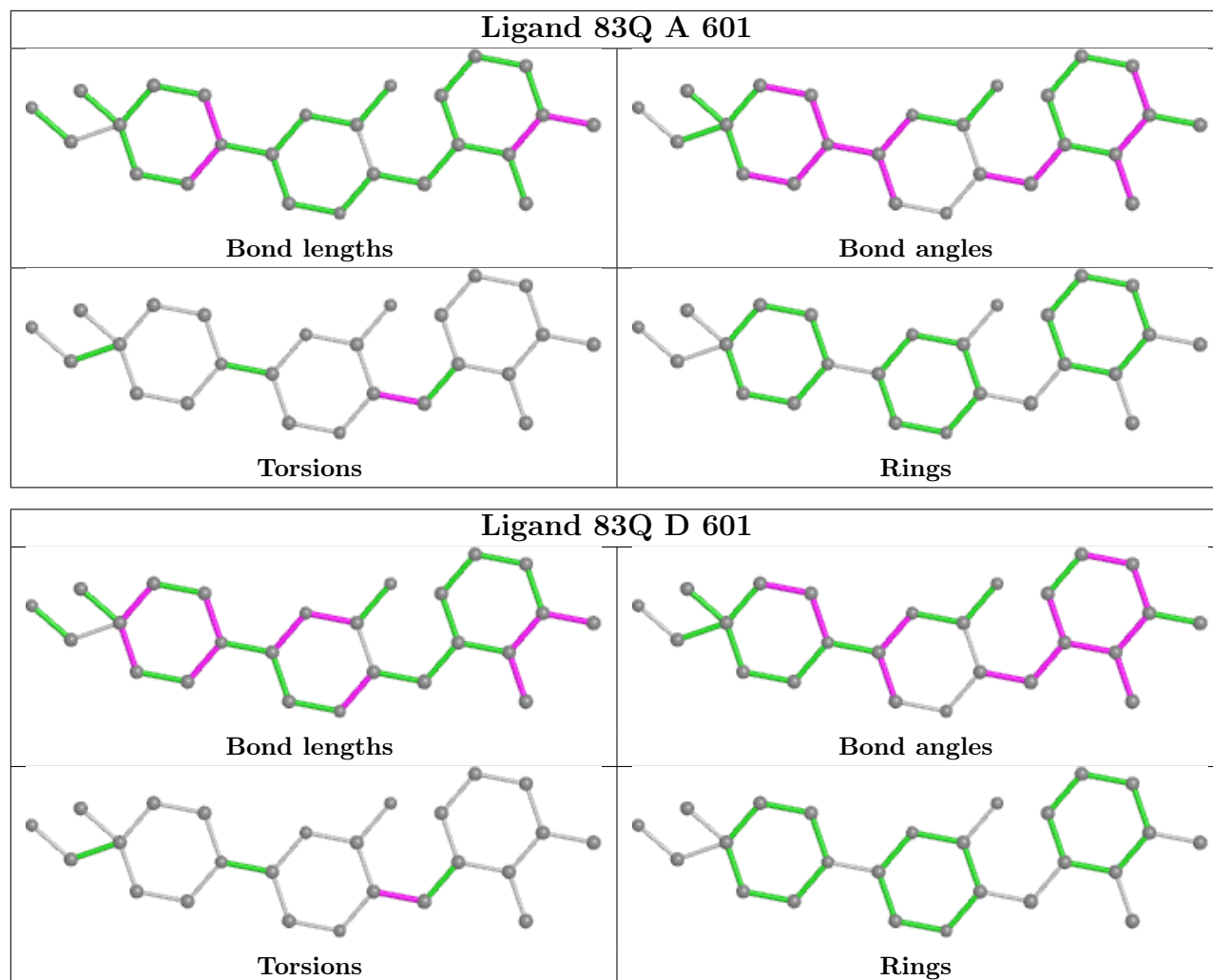
*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	601	83Q	2	0
2	D	601	83Q	1	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less than 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.





## 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	477/526 (90%)	0.64	26 (5%)	25	31	17, 35, 62, 84	0
1	B	480/526 (91%)	0.65	28 (5%)	23	28	19, 35, 61, 94	0
1	C	481/526 (91%)	0.64	33 (6%)	16	21	19, 34, 62, 107	0
1	D	487/526 (92%)	0.68	36 (7%)	14	18	17, 35, 64, 117	0
All	All	1925/2104 (91%)	0.65	123 (6%)	19	24	17, 35, 62, 117	0

The worst 5 of 123 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	36	SER	6.8
1	B	36	SER	5.5
1	B	95	VAL	4.9
1	D	484	VAL	4.5
1	D	38	PRO	4.4

### 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](#)

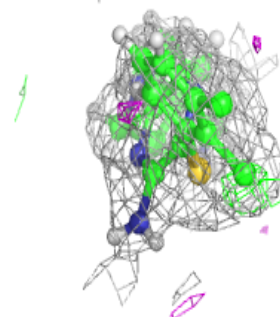
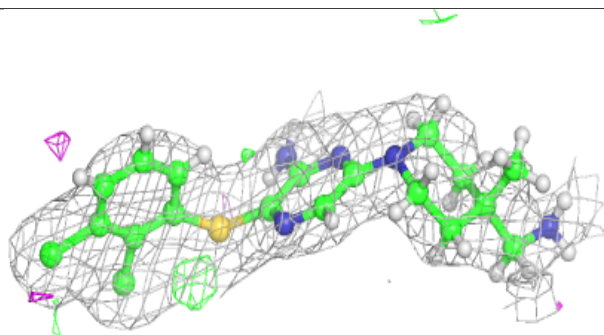
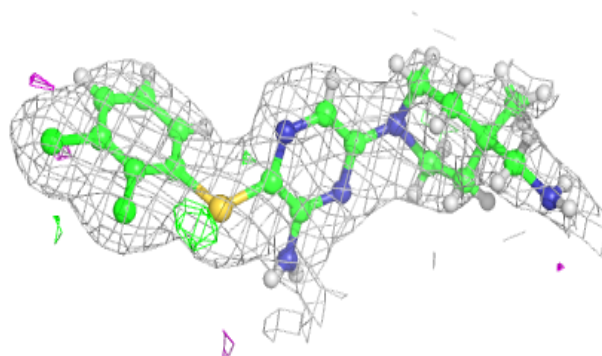
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled ‘Q < 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
2	83Q	D	601	25/25	0.92	0.15	19,32,45,47	0
2	83Q	B	601	25/25	0.94	0.12	11,31,49,53	0
2	83Q	C	601	25/25	0.94	0.14	16,30,38,49	0
2	83Q	A	601	25/25	0.94	0.14	14,31,42,49	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

**Electron density around 83Q D 601:**

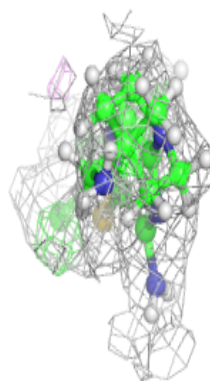
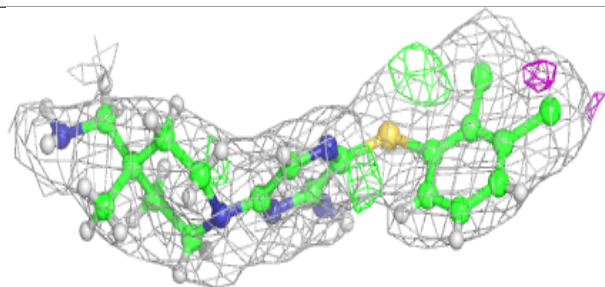
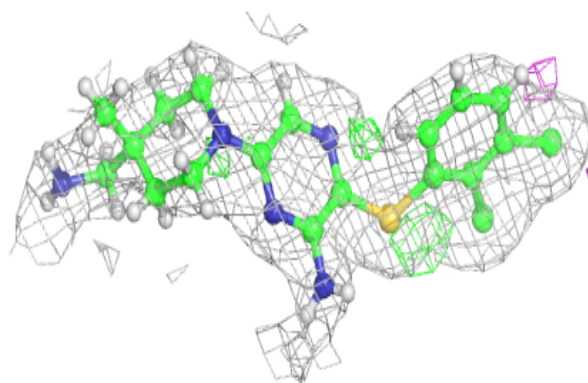
2mF<sub>o</sub>-DF<sub>c</sub> (at 0.7 rmsd) in gray  
mF<sub>o</sub>-DF<sub>c</sub> (at 3 rmsd) in purple (negative)  
and green (positive)



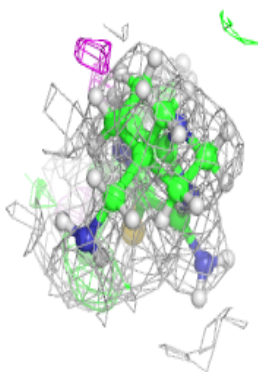
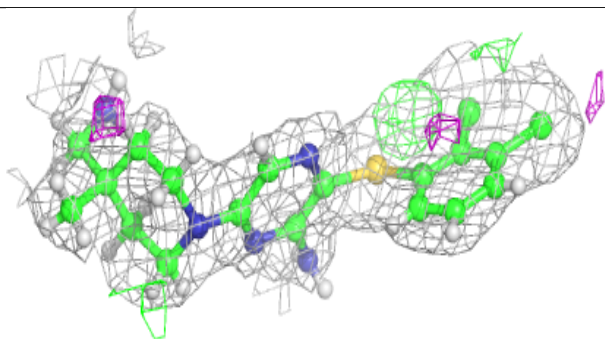
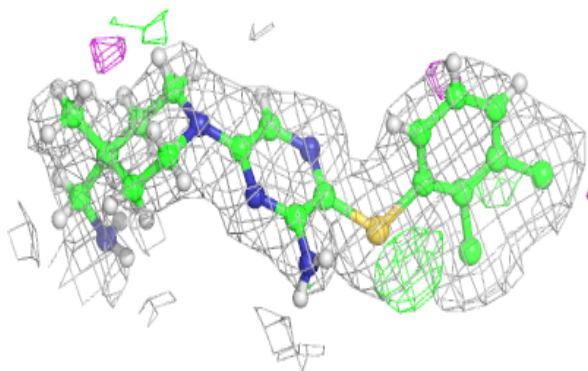


**Electron density around 83Q B 601:**

$2mF_o-DF_c$  (at 0.7 rmsd) in gray  
 $mF_o-DF_c$  (at 3 rmsd) in purple (negative)  
and green (positive)

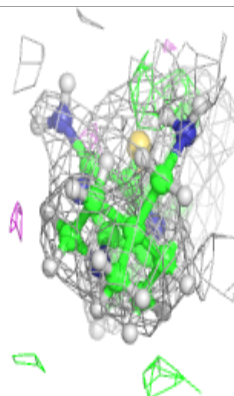
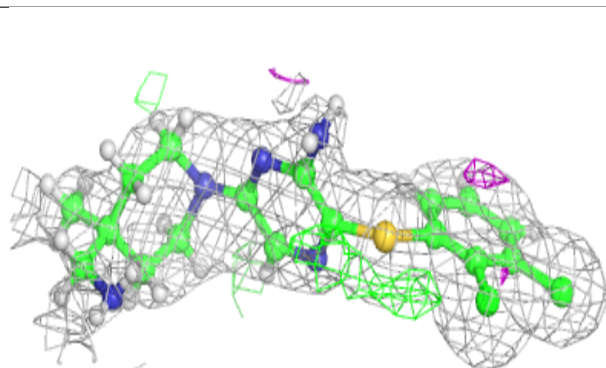
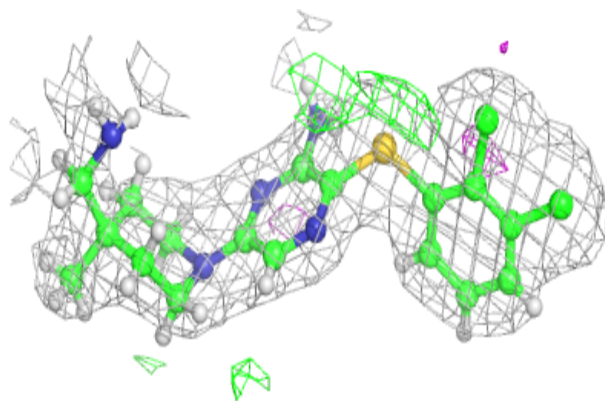
**Electron density around 83Q C 601:**

$2mF_o-DF_c$  (at 0.7 rmsd) in gray  
 $mF_o-DF_c$  (at 3 rmsd) in purple (negative)  
and green (positive)



**Electron density around 83Q A 601:**

$2mF_o-DF_c$  (at 0.7 rmsd) in gray  
 $mF_o-DF_c$  (at 3 rmsd) in purple (negative)  
and green (positive)



## 6.5 Other polymers [i](#)

There are no such residues in this entry.