



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 5, 2026 – 03:02 PM UTC

PDB ID : 9YG1 / pdb_00009yg1
Title : Targeting PTPN22 at non-orthosteric binding sites - a fragment approach
Authors : Di Lello, P.; Wells, M.M.
Deposited on : 2025-09-27
Resolution : 1.97 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

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity : 4-5-2 with Phenix2.0
Mogul : 2022.3.0, CSD as543be (2022)
Xtriage (Phenix) : 2.0
EDS : 3.0
Buster-report : wwPDB partial adaption of 1.1.7 (2018)
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4 : 9.0.010 (Gargrove)
Density-Fitness : 1.0.12
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.49

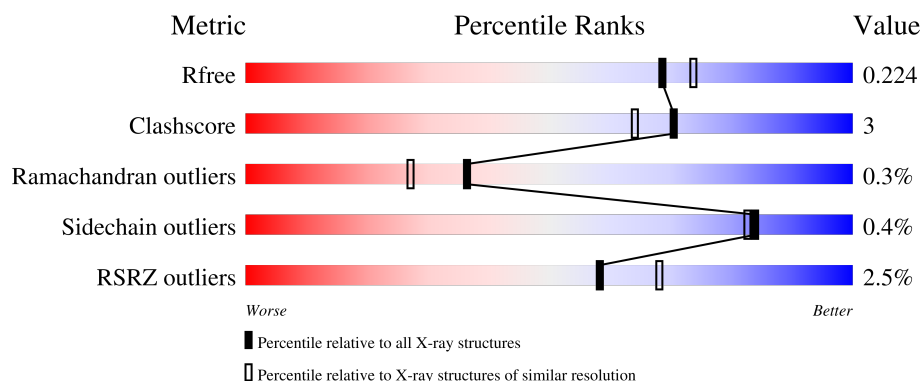
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.97 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}	180053	1506 (1.98-1.98)
Clashscore	190562	1534 (1.98-1.98)
Ramachandran outliers	187476	1518 (1.98-1.98)
Sidechain outliers	187428	1518 (1.98-1.98)
RSRZ outliers	180081	1506 (1.98-1.98)

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 $\leq 5\%$. 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	322	3% 86% 9% 5%
1	B	322	2% 87% 6% 7%

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
2	EOH	A	402	-	-	X	-

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 10284 atoms, of which 4859 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 22.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	307	Total	C	H	N	O	S	0	1	0
			4890	1597	2387	414	475	17			
1	B	300	Total	C	H	N	O	S	0	2	0
			4888	1587	2414	406	462	19			

There are 38 discrepancies between the modelled and reference sequences:

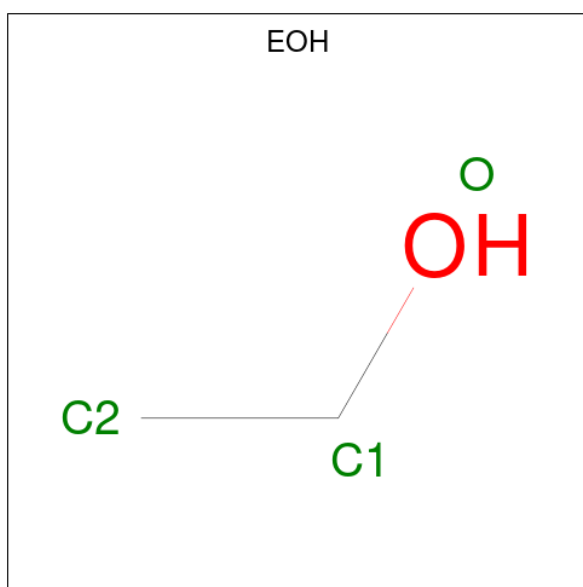
Chain	Residue	Modelled	Actual	Comment	Reference
A	-15	MET	-	initiating methionine	UNP Q9Y2R2
A	-14	HIS	-	expression tag	UNP Q9Y2R2
A	-13	HIS	-	expression tag	UNP Q9Y2R2
A	-12	HIS	-	expression tag	UNP Q9Y2R2
A	-11	HIS	-	expression tag	UNP Q9Y2R2
A	-10	HIS	-	expression tag	UNP Q9Y2R2
A	-9	HIS	-	expression tag	UNP Q9Y2R2
A	-8	GLY	-	expression tag	UNP Q9Y2R2
A	-7	GLU	-	expression tag	UNP Q9Y2R2
A	-6	ASN	-	expression tag	UNP Q9Y2R2
A	-5	LEU	-	expression tag	UNP Q9Y2R2
A	-4	TYR	-	expression tag	UNP Q9Y2R2
A	-3	PHE	-	expression tag	UNP Q9Y2R2
A	-2	GLN	-	expression tag	UNP Q9Y2R2
A	-1	GLY	-	expression tag	UNP Q9Y2R2
A	0	SER	-	expression tag	UNP Q9Y2R2
A	304	GLY	-	expression tag	UNP Q9Y2R2
A	305	ASN	-	expression tag	UNP Q9Y2R2
A	306	SER	-	expression tag	UNP Q9Y2R2
B	-15	MET	-	initiating methionine	UNP Q9Y2R2
B	-14	HIS	-	expression tag	UNP Q9Y2R2
B	-13	HIS	-	expression tag	UNP Q9Y2R2
B	-12	HIS	-	expression tag	UNP Q9Y2R2
B	-11	HIS	-	expression tag	UNP Q9Y2R2
B	-10	HIS	-	expression tag	UNP Q9Y2R2

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-9	HIS	-	expression tag	UNP Q9Y2R2
B	-8	GLY	-	expression tag	UNP Q9Y2R2
B	-7	GLU	-	expression tag	UNP Q9Y2R2
B	-6	ASN	-	expression tag	UNP Q9Y2R2
B	-5	LEU	-	expression tag	UNP Q9Y2R2
B	-4	TYR	-	expression tag	UNP Q9Y2R2
B	-3	PHE	-	expression tag	UNP Q9Y2R2
B	-2	GLN	-	expression tag	UNP Q9Y2R2
B	-1	GLY	-	expression tag	UNP Q9Y2R2
B	0	SER	-	expression tag	UNP Q9Y2R2
B	304	GLY	-	expression tag	UNP Q9Y2R2
B	305	ASN	-	expression tag	UNP Q9Y2R2
B	306	SER	-	expression tag	UNP Q9Y2R2

- Molecule 2 is ETHANOL (CCD ID: EOH) (formula: C₂H₆O).



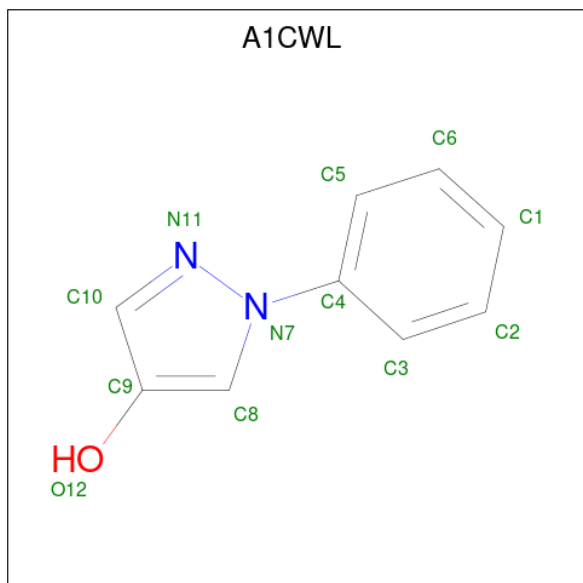
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	H	O	0	0
			9	2	6	1		
2	A	1	Total	C	H	O	0	0
			9	2	6	1		
2	A	1	Total	C	H	O	0	0
			9	2	6	1		
2	B	1	Total	C	H	O	0	0
			9	2	6	1		
2	B	1	Total	C	H	O	0	0
			9	2	6	1		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	B	1	Total	C	H	O	0	0
			9	2	6	1		
2	B	1	Total	C	H	O	0	0
			9	2	6	1		

- Molecule 3 is 1-phenyl-1H-pyrazol-4-ol (CCD ID: A1CWL) (formula: C₉H₈N₂O) (labeled as "Ligand of Interest" by depositor).



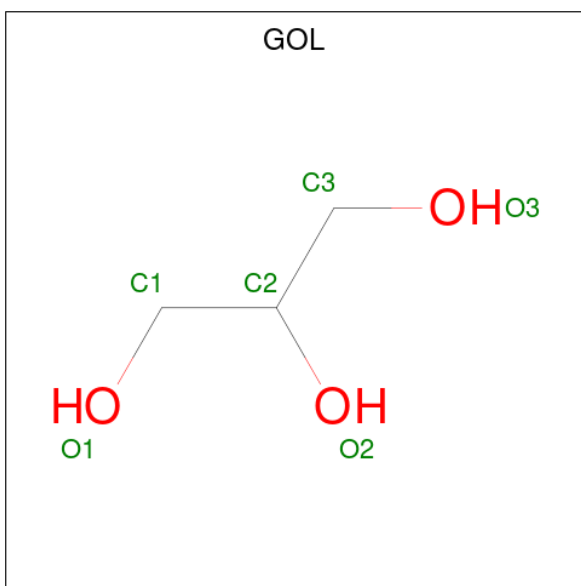
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	A	1	Total	C	H	N	O	0	0
			20	9	8	2	1		

- Molecule 4 is SULFATE ION (CCD ID: SO4) (formula: O₄S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	B	1	Total	O	S	0	0
			5	4	1		

- Molecule 5 is GLYCEROL (CCD ID: GOL) (formula: C₃H₈O₃).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	B	1	Total	C	H	O	0	0
			14	3	8	3		

- Molecule 6 is MAGNESIUM ION (CCD ID: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	B	2	Total 2	Mg 2	0	0

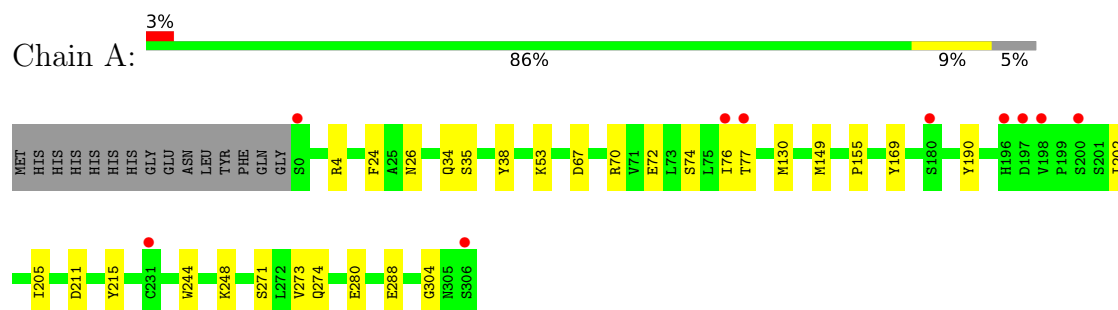
- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	205	Total 205	O 205	0	0
7	B	197	Total 197	O 197	0	0

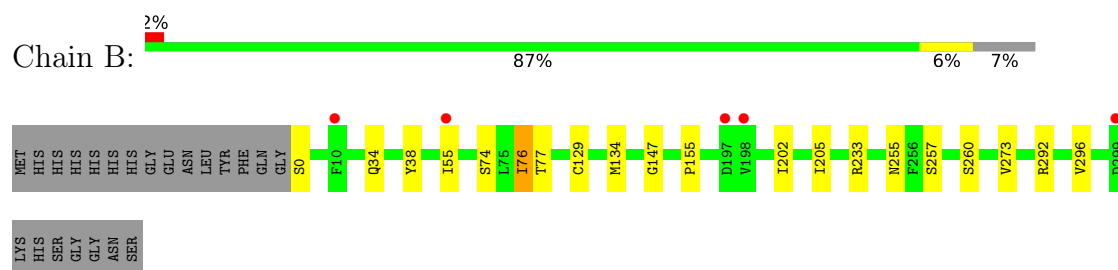
3 Residue-property plots [i](#)

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: Tyrosine-protein phosphatase non-receptor type 22



- Molecule 1: Tyrosine-protein phosphatase non-receptor type 22



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	60.14Å 48.16Å 119.71Å 90.00° 102.50° 90.00°	Depositor
Resolution (Å)	116.88 – 1.97 116.88 – 1.97	Depositor EDS
% Data completeness (in resolution range)	91.8 (116.88-1.97) 91.7 (116.88-1.97)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.65 (at 1.88Å)	Xtriage
Refinement program	PHENIX (1.20.1_4487: ???)	Depositor
R, R_{free}	0.176 , 0.224 0.176 , 0.224	Depositor DCC
R_{free} test set	2281 reflections (4.18%)	wwPDB-VP
Wilson B-factor (Å ²)	35.2	Xtriage
Anisotropy	0.029	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.40 , 40.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.013 for h,-k,-h-l	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	10284	wwPDB-VP
Average B, all atoms (Å ²)	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 11.19% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: EOH, MG, A1CWL, SO4, GOL

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	0.31	0/2561	0.47	0/3465
1	B	0.32	0/2537	0.48	0/3428
All	All	0.31	0/5098	0.48	0/6893

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	A	2503	2387	2428	22	0
1	B	2474	2414	2448	12	0
2	A	9	18	18	5	0
2	B	12	24	24	0	0
3	A	12	8	0	0	0
4	B	5	0	0	0	0
5	B	6	8	8	1	0
6	B	2	0	0	0	0
7	A	205	0	0	4	0
7	B	197	0	0	2	0
All	All	5425	4859	4926	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 3.

The worst 5 of 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:B:129[A]:CYS:SG	1:B:233:ARG:NH2	2.60	0.74
1:A:74:SER:O	1:A:76:ILE:HD12	1.96	0.66
1:B:129[A]:CYS:SG	1:B:233:ARG:CZ	2.89	0.60
1:A:26:ASN:ND2	7:A:502:HOH:O	2.36	0.58
1:A:304:GLY:N	2:A:402:EOH:H23	2.22	0.55

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	306/322 (95%)	296 (97%)	9 (3%)	1 (0%)	36	27
1	B	300/322 (93%)	290 (97%)	9 (3%)	1 (0%)	36	27
All	All	606/644 (94%)	586 (97%)	18 (3%)	2 (0%)	36	27

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	273	VAL
1	B	273	VAL

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	273/295 (92%)	273 (100%)	0	100	100
1	B	274/295 (93%)	272 (99%)	2 (1%)	76	73
All	All	547/590 (93%)	545 (100%)	2 (0%)	84	82

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	55	ILE
1	B	76	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	8	GLN
1	A	26	ASN
1	B	150	GLN

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 oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 12 ligands modelled in this entry, 2 are monoatomic - leaving 10 for Mogul analysis.

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	EOH	B	403	-	2,2,2	0.45	0	1,1,1	0.33	0
2	EOH	A	401	-	2,2,2	0.50	0	1,1,1	0.04	0
2	EOH	A	402	-	2,2,2	0.51	0	1,1,1	0.07	0
4	SO4	B	401	-	4,4,4	0.32	0	6,6,6	0.15	0
5	GOL	B	402	-	5,5,5	0.91	0	5,5,5	0.91	0
2	EOH	B	405	-	2,2,2	0.44	0	1,1,1	0.07	0
2	EOH	A	403	-	2,2,2	0.46	0	1,1,1	0.20	0
2	EOH	B	404	-	2,2,2	0.45	0	1,1,1	0.31	0
2	EOH	B	406	-	2,2,2	0.48	0	1,1,1	0.14	0
3	A1CWL	A	404	-	13,13,13	1.63	2 (15%)	13,17,17	2.22	4 (30%)

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
5	GOL	B	402	-	-	0/4/4/4	-
3	A1CWL	A	404	-	-	0/4/4/4	0/2/2/2

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	404	A1CWL	C4-N7	-5.06	1.33	1.43
3	A	404	A1CWL	O12-C9	2.13	1.39	1.33

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	404	A1CWL	C10-N11-N7	5.69	108.72	104.19
3	A	404	A1CWL	C4-N7-C8	-3.20	124.49	129.12
3	A	404	A1CWL	C4-N7-N11	3.00	126.21	121.08
3	A	404	A1CWL	C8-N7-N11	-2.79	109.15	112.14

There are no chirality outliers.

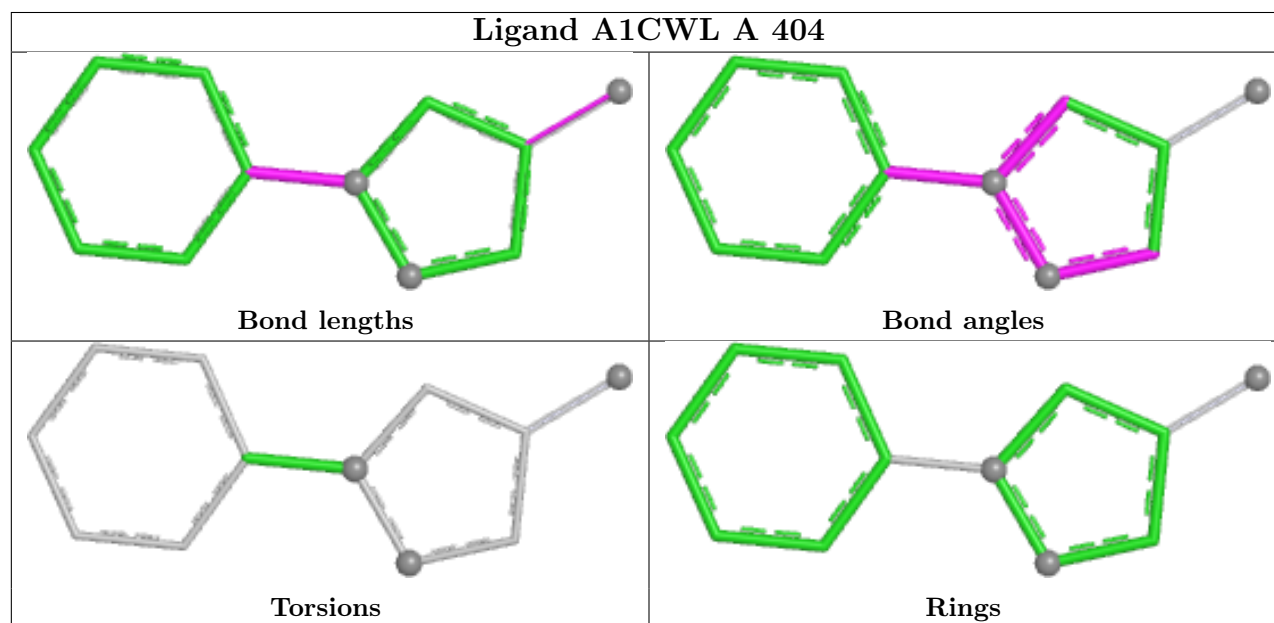
There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	401	EOH	1	0
2	A	402	EOH	4	0
5	B	402	GOL	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, 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	307/322 (95%)	0.06	10 (3%) 49 59	17, 42, 72, 99	1 (0%)
1	B	300/322 (93%)	0.02	5 (1%) 69 78	25, 42, 71, 104	2 (0%)
All	All	607/644 (94%)	0.04	15 (2%) 58 68	17, 42, 71, 104	3 (0%)

The worst 5 of 15 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	197	ASP	5.0
1	A	197	ASP	3.7
1	B	299	ASP	3.4
1	A	0	SER	3.3
1	A	76	ILE	3.0

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 oligosaccharides in this entry.

6.4 Ligands [i](#)

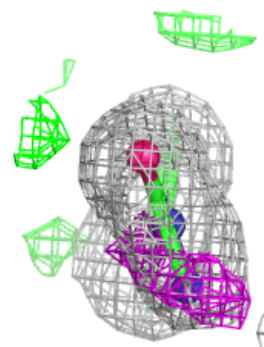
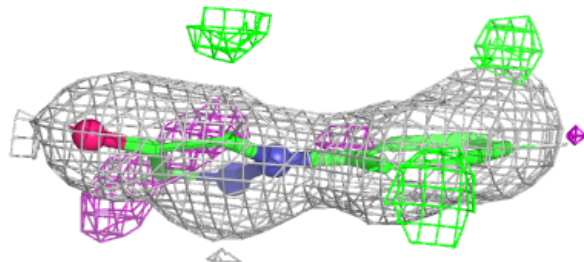
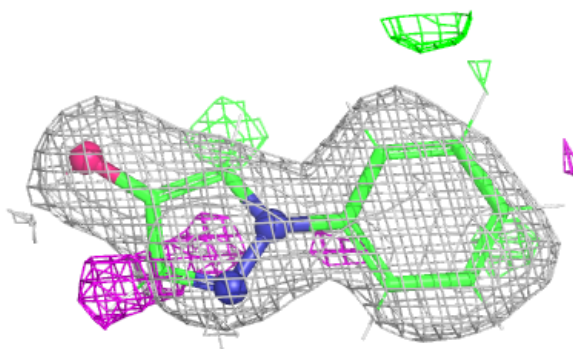
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled ‘Q< 0.9’ lists 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	EOH	B	406	3/3	0.60	0.25	54,65,72,72	0
2	EOH	B	404	3/3	0.76	0.19	42,50,62,71	0
2	EOH	B	403	3/3	0.76	0.21	56,68,74,74	0
2	EOH	A	403	3/3	0.77	0.20	58,70,76,76	0
2	EOH	A	402	3/3	0.80	0.18	44,52,64,72	0
4	SO4	B	401	5/5	0.81	0.13	51,52,57,69	0
2	EOH	A	401	3/3	0.82	0.13	39,47,58,66	0
2	EOH	B	405	3/3	0.84	0.18	45,54,60,61	0
5	GOL	B	402	6/6	0.90	0.10	44,52,61,62	0
3	A1CWL	A	404	12/12	0.91	0.10	26,39,47,48	0
6	MG	B	408	1/1	0.97	0.12	35,35,35,35	0
6	MG	B	407	1/1	1.00	0.06	30,30,30,30	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 A1CWL A 404:

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.