Full wwPDB X-ray Structure Validation Report (i)

Nov 3, 2026 – 06:07 pm GMT

PDB ID	:	32TH / pdb_000032th
Title	:	Crystal structure of YTHDF2 in complex with compound 20 (ZA801)
Authors	:	Nai, F.; Zalesak, F.; Caflisch, A.
Deposited on	:	2026-07-23
Resolution	:	1.68 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:

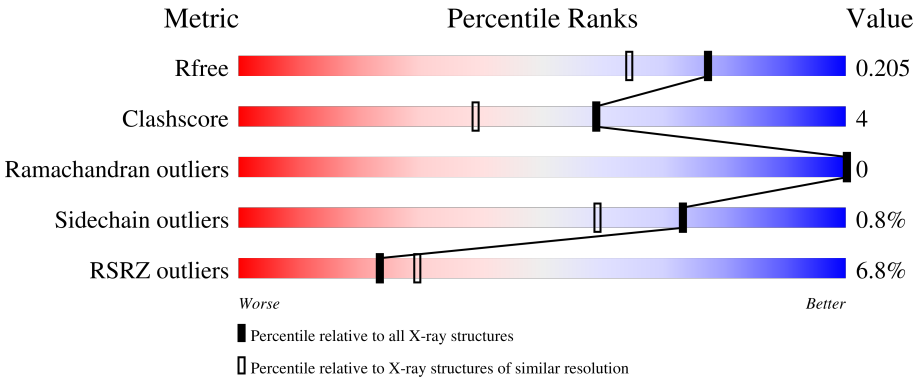
MolProbity	:	4-5-2 with Phenix2.2
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	2.2
EDS	:	3.0
Buster-report	:	wwPDB partial adaption of 1.1.7 (2018)
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.015 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.51

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 1.68 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	1054 (1.68-1.68)
Clashscore	190562	1078 (1.68-1.68)
Ramachandran outliers	187476	1068 (1.68-1.68)
Sidechain outliers	187428	1067 (1.68-1.68)
RSRZ outliers	180081	1055 (1.68-1.68)

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	167	6% 80% 7% • 13%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 1404 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.

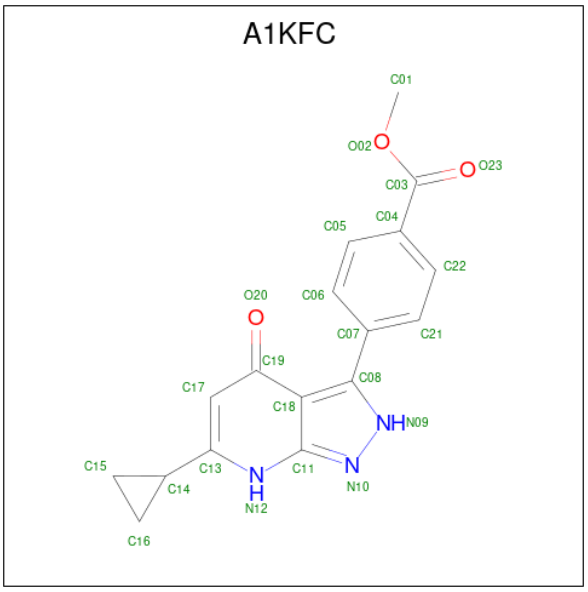
- Molecule 1 is a protein called YTH domain-containing family protein 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	146	Total	C	N	O	S	0	6	0
			1224	774	221	223	6			

There are 22 discrepancies between the modelled and reference sequences:

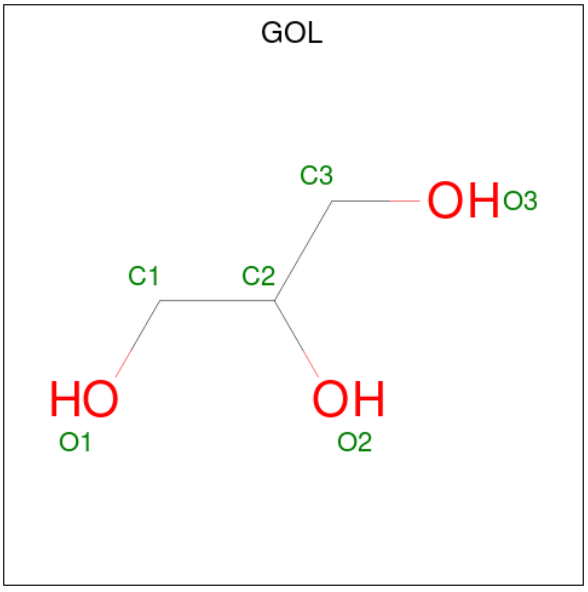
Chain	Residue	Modelled	Actual	Comment	Reference
A	386	MET	-	initiating methionine	UNP Q9Y5A9
A	387	GLY	-	expression tag	UNP Q9Y5A9
A	388	SER	-	expression tag	UNP Q9Y5A9
A	389	SER	-	expression tag	UNP Q9Y5A9
A	390	TYR	-	expression tag	UNP Q9Y5A9
A	391	HIS	-	expression tag	UNP Q9Y5A9
A	392	HIS	-	expression tag	UNP Q9Y5A9
A	393	HIS	-	expression tag	UNP Q9Y5A9
A	394	HIS	-	expression tag	UNP Q9Y5A9
A	395	HIS	-	expression tag	UNP Q9Y5A9
A	396	HIS	-	expression tag	UNP Q9Y5A9
A	397	SER	-	expression tag	UNP Q9Y5A9
A	398	SER	-	expression tag	UNP Q9Y5A9
A	399	GLY	-	expression tag	UNP Q9Y5A9
A	400	GLU	-	expression tag	UNP Q9Y5A9
A	401	ASN	-	expression tag	UNP Q9Y5A9
A	402	LEU	-	expression tag	UNP Q9Y5A9
A	403	TYR	-	expression tag	UNP Q9Y5A9
A	404	PHE	-	expression tag	UNP Q9Y5A9
A	405	GLN	-	expression tag	UNP Q9Y5A9
A	406	HIS	-	expression tag	UNP Q9Y5A9
A	407	MET	-	expression tag	UNP Q9Y5A9

- Molecule 2 is methyl 4-(6-cyclopropyl-4-oxidanylidene-pyrazolo[3,4-b]pyridin-3-yl)benzoate (CCD ID: A1KFC) (formula: $C_{17}H_{15}N_3O_3$).



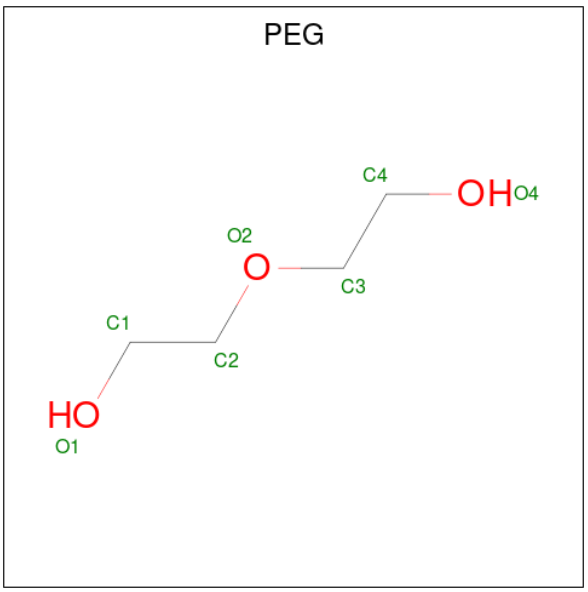
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	0	0
			23	17	3	3		

- Molecule 3 is GLYCEROL (CCD ID: GOL) (formula: $C_3H_8O_3$).



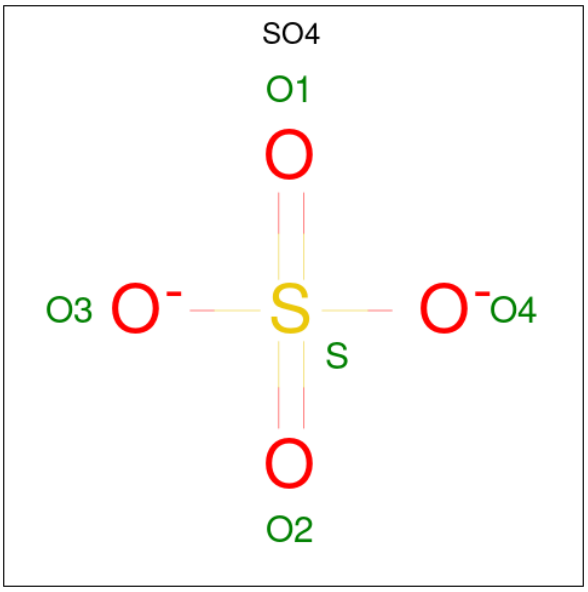
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	O		0	0
			6	3	3			
3	A	1	Total	C	O		0	0
			6	3	3			
3	A	1	Total	C	O		0	0
			6	3	3			

- Molecule 4 is DI(HYDROXYETHYL)ETHER (CCD ID: PEG) (formula: $C_4H_{10}O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C O 7 4 3	0	0
4	A	1	Total C O 7 4 3	0	0

- Molecule 5 is SULFATE ION (CCD ID: SO4) (formula: O_4S).



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

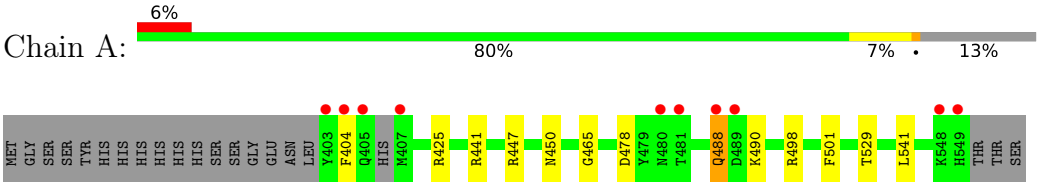
- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	120	Total O 120 120	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: YTH domain-containing family protein 2



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	43.80Å 46.30Å 72.84Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	37.53 – 1.68 37.53 – 1.68	Depositor EDS
% Data completeness (in resolution range)	99.8 (37.53-1.68) 99.8 (37.53-1.68)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.37 (at 1.68Å)	Xtriage
Refinement program	PHENIX 2.1_6048	Depositor
R, R_{free}	0.182 , 0.201 0.184 , 0.205	Depositor DCC
R_{free} test set	871 reflections (4.94%)	wwPDB-VP
Wilson B-factor (Å^2)	17.8	Xtriage
Anisotropy	0.803	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 40.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.52$, $< L^2 > = 0.36$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	1404	wwPDB-VP
Average B, all atoms (Å^2)	22.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 11.00% of the height of the origin peak. No significant pseudotranslation is detected.*

[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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, PEG, SO4, A1KFC

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	0.25	0/1253	0.44	0/1690

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	1224	0	1164	10	0
2	A	23	0	0	0	0
3	A	18	0	24	3	0
4	A	14	0	20	0	0
5	A	5	0	0	0	0
6	A	120	0	0	3	0
All	All	1404	0	1208	11	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 (11) 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:498:ARG:NE	6:A:701:HOH:O	2.27	0.66
1:A:465:GLY:HA2	3:A:602:GOL:H2	1.85	0.58
1:A:450[B]:ASN:ND2	1:A:450[B]:ASN:O	2.38	0.56
1:A:441[B]:ARG:NH1	1:A:529:THR:O	2.42	0.53
1:A:478:ASP:HB3	6:A:730:HOH:O	2.09	0.52
1:A:425:ARG:HE	3:A:603:GOL:H12	1.75	0.52
1:A:488[A]:GLN:HG2	1:A:490:LYS:H	1.79	0.47
1:A:447:ARG:HA	1:A:450[B]:ASN:OD1	2.14	0.47
1:A:404:PHE:CE2	1:A:501:PHE:HE2	2.37	0.42
3:A:604:GOL:H11	6:A:807:HOH:O	2.20	0.41
1:A:404:PHE:CZ	1:A:541:LEU:HD22	2.56	0.41

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	148/167 (89%)	146 (99%)	2 (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	129/149 (87%)	127 (98%)	2 (2%)	55 32

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

Mol	Chain	Res	Type
1	A	488[A]	GLN
1	A	488[B]	GLN

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	409	HIS
1	A	509	GLN
1	A	530	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)

7 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	$\#\|Z\| > 2$
4	PEG	A	606	-	6,6,6	0.26	0	5,5,5	0.23	0
5	SO4	A	607	-	4,4,4	0.51	0	6,6,6	0.27	0
3	GOL	A	604	-	5,5,5	0.36	0	5,5,5	0.43	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\|>2$	Counts	RMSZ	$\#\|Z\|>2$
2	A1KFC	A	601	-	26,26,26	1.64	3 (11%)	28,38,38	2.71	7 (25%)
3	GOL	A	603	-	5,5,5	0.35	0	5,5,5	0.37	0
4	PEG	A	605	-	6,6,6	0.22	0	5,5,5	0.23	0
3	GOL	A	602	-	5,5,5	0.29	0	5,5,5	0.55	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
4	PEG	A	606	-	-	2/4/4/4	-
3	GOL	A	604	-	-	0/4/4/4	-
2	A1KFC	A	601	-	-	2/12/16/16	0/4/4/4
3	GOL	A	603	-	-	4/4/4/4	-
4	PEG	A	605	-	-	4/4/4/4	-
3	GOL	A	602	-	-	2/4/4/4	-

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	601	A1KFC	C07-C08	-3.70	1.42	1.47
2	A	601	A1KFC	O02-C03	3.36	1.40	1.33
2	A	601	A1KFC	C18-C08	3.32	1.45	1.39

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	601	A1KFC	C11-N10-N09	8.54	109.33	103.24
2	A	601	A1KFC	N12-C11-N10	6.36	128.69	123.31
2	A	601	A1KFC	C21-C07-C08	-4.67	115.17	120.76
2	A	601	A1KFC	C14-C13-C17	-4.43	119.62	125.36
2	A	601	A1KFC	C06-C07-C08	3.58	125.05	120.76
2	A	601	A1KFC	C18-C11-N10	-3.36	108.88	113.63
2	A	601	A1KFC	O02-C03-C04	3.15	117.12	112.34

There are no chirality outliers.

All (14) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	601	A1KFC	C17-C13-C14-C15
2	A	601	A1KFC	C17-C13-C14-C16
3	A	602	GOL	C1-C2-C3-O3
3	A	603	GOL	O2-C2-C3-O3
4	A	605	PEG	O1-C1-C2-O2
3	A	603	GOL	O1-C1-C2-C3
3	A	603	GOL	C1-C2-C3-O3
3	A	602	GOL	O2-C2-C3-O3
4	A	606	PEG	O1-C1-C2-O2
4	A	605	PEG	C4-C3-O2-C2
3	A	603	GOL	O1-C1-C2-O2
4	A	605	PEG	O2-C3-C4-O4
4	A	606	PEG	C4-C3-O2-C2
4	A	605	PEG	C1-C2-O2-C3

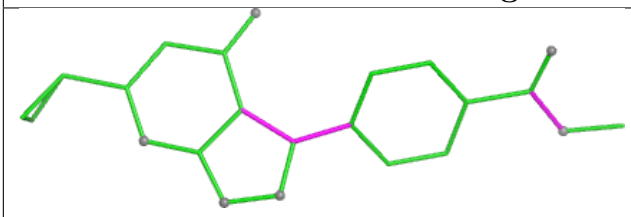
There are no ring outliers.

3 monomers are involved in 3 short contacts:

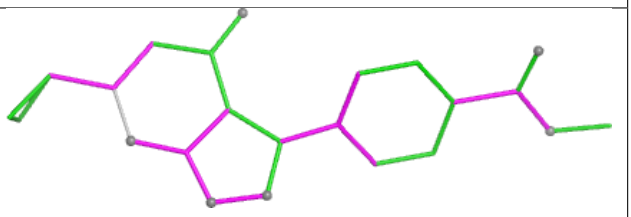
Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	604	GOL	1	0
3	A	603	GOL	1	0
3	A	602	GOL	1	0

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 then 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.

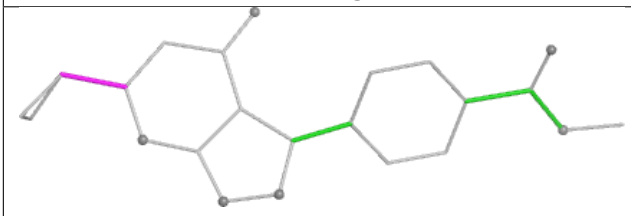
Ligand A1KFC A 601



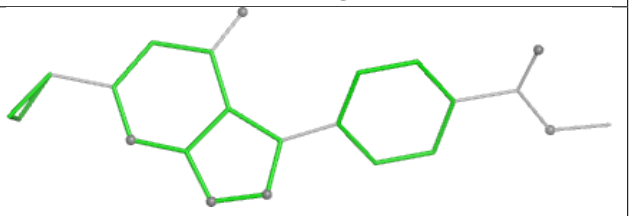
Bond lengths



Bond angles



Torsions



Rings

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

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^{th} 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	146/167 (87%)	0.19	10 (6%) 23 29	9, 18, 37, 59	6 (4%)

All (10) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	404	PHE	5.8
1	A	405	GLN	4.7
1	A	403	TYR	4.7
1	A	549	HIS	4.2
1	A	480	ASN	3.5
1	A	488[A]	GLN	3.3
1	A	489	ASP	2.5
1	A	548	LYS	2.5
1	A	407	MET	2.4
1	A	481	THR	2.2

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no oligosaccharides in this entry.

6.4 Ligands

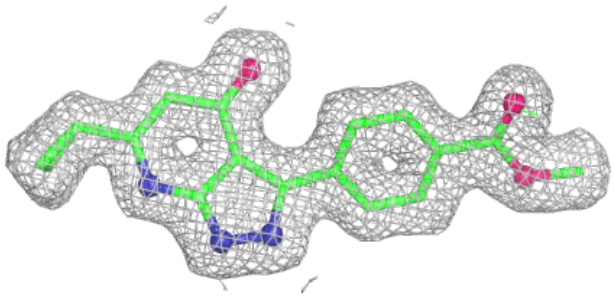
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

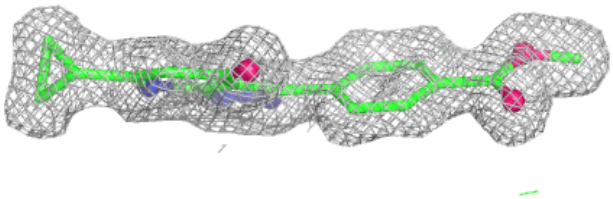
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	GOL	A	604	6/6	0.66	0.20	40,44,57,59	0
3	GOL	A	603	6/6	0.73	0.17	37,44,52,62	0
4	PEG	A	606	7/7	0.73	0.16	32,38,44,44	0
3	GOL	A	602	6/6	0.76	0.17	30,39,42,47	0
4	PEG	A	605	7/7	0.79	0.16	32,41,48,53	0
2	A1KFC	A	601	23/23	0.96	0.05	12,15,17,17	0
5	SO4	A	607	5/5	0.98	0.05	18,18,21,23	0

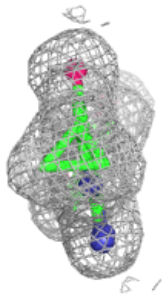
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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

Electron density around A1KFC 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

There are no such residues in this entry.