wwPDB X-ray Structure Validation Summary Report (i)

Jun 5, 2023 – 03:03 PM EDT

PDB ID	:	8G3I
Title	:	Non-ribosomal PCP-C didomain (thioether stabilised glycolic acid) acceptor bound state
Authors	:	Ho, Y.T.C.; Izore, T.; Cryle, M.J.
Deposited on	:	2023-02-08
Resolution	:	3.02 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 (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:

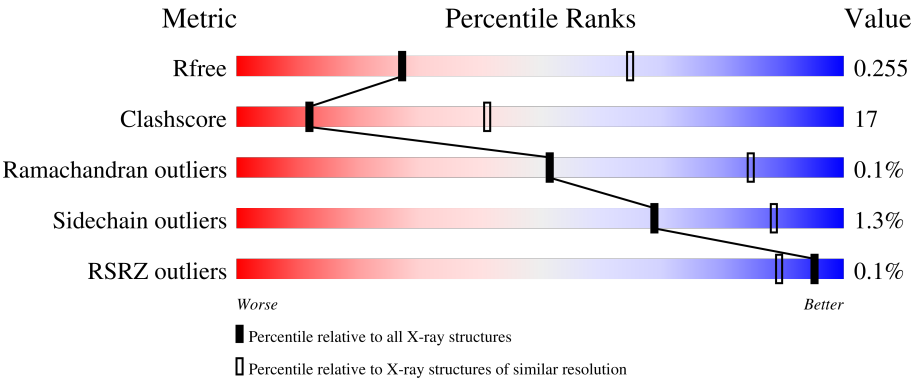
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.33
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.33

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 3.02 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	2399 (3.04-3.00)
Clashscore	141614	2734 (3.04-3.00)
Ramachandran outliers	138981	2640 (3.04-3.00)
Sidechain outliers	138945	2643 (3.04-3.00)
RSRZ outliers	127900	2287 (3.04-3.00)

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	528	63% 33%
1	B	528	65% 32%

2 Entry composition (i)

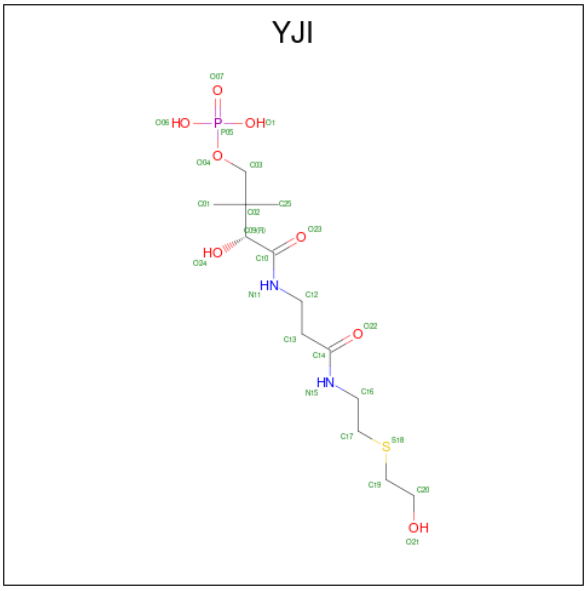
There are 3 unique types of molecules in this entry. The entry contains 8011 atoms, of which 50 are hydrogens and 0 are deuteriums.

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 PCP-C didomain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	509	Total	C	N	O	S	0	0	0
			3933	2485	711	729	8			
1	B	515	Total	C	N	O	S	0	0	0
			3971	2506	720	737	8			

- Molecule 2 is N 3 -[(2R)-2-hydroxy-3,3-dimethyl-4-(phosphonooxy)butanoyl]-N-{2-[(2-hydroxyethyl)sulfanyl]ethyl}-beta-alaninamide (three-letter code: YJI) (formula: $C_{13}H_{27}N_2O_8PS$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms							ZeroOcc	AltConf
2	A	1	Total	C	H	N	O	P	S	0	0
			49	13	25	2	7	1	1		
2	B	1	Total	C	H	N	O	P	S	0	0
			49	13	25	2	7	1	1		

- Molecule 3 is water.

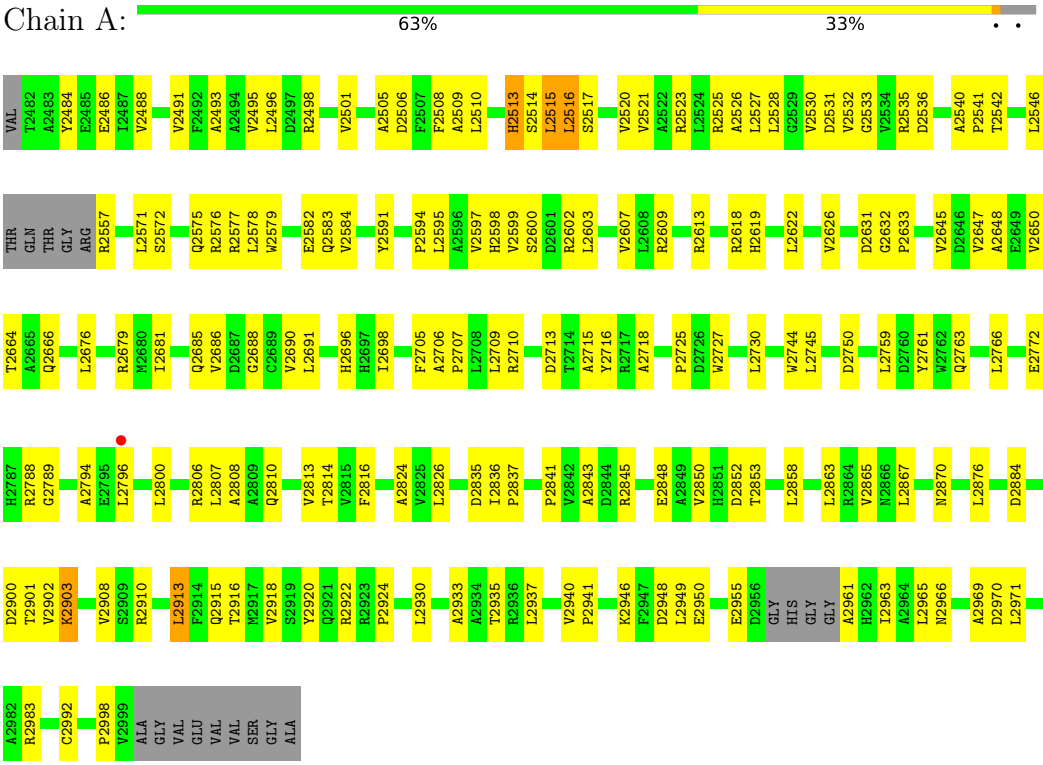
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	4	Total O 4 4	0	0
3	B	5	Total O 5 5	0	0

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.

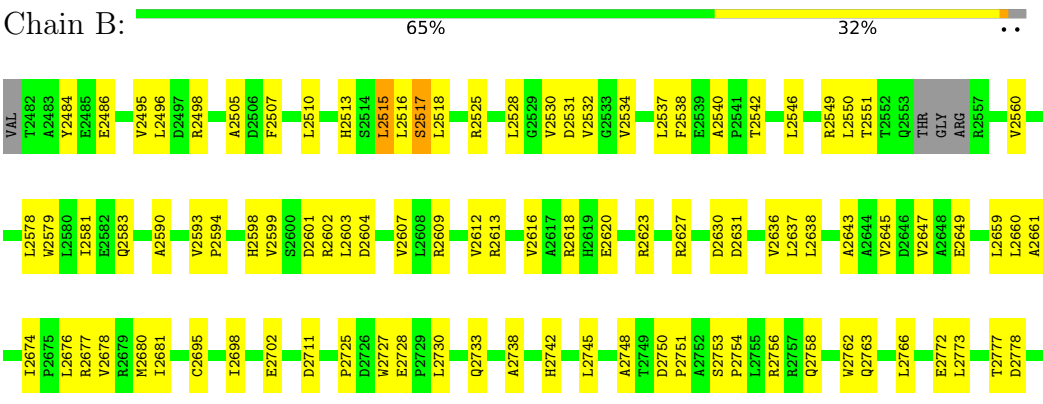
• Molecule 1: PCP-C didomain

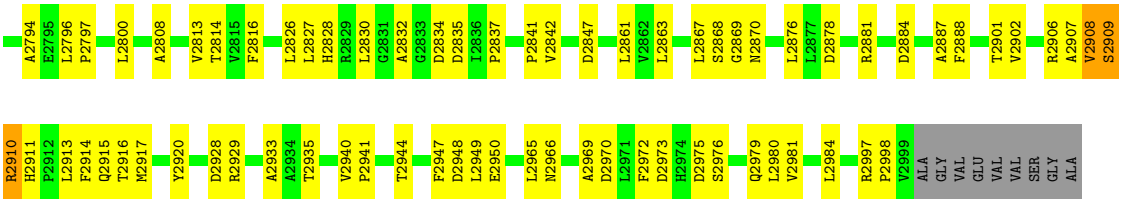
Chain A:



• Molecule 1: PCP-C didomain

Chain B:




A2794 E2795 L2796 P2797 L2800 A2808 V2813 T2814 V2815 F2816 L2826 L2827 H2828 R2829 L2830 G2831 A2832 G2833 D2834 D2835 I2836 P2837 P2841 V2842 D2847 L2861 V2862 L2863 L2867 S2868 G2869 N2870 L2876 L2877 D2878 R2881 D2884 A2887 F2888 T2901 V2902 R2906 A2907 V2908 S2909
R2910 H2911 P2912 L2913 F2914 Q2915 T2916 M2917 Y2920 D2928 R2929 A2933 A2934 T2935 V2940 P2941 T2944 F2947 D2948 L2949 E2950 L2965 N2966 A2969 D2970 L2971 F2972 D2973 H2974 D2975 S2976 Q2979 L2980 V2981 L2984 R2997 P2998 V2999 ALA GLY VAL GLU VAL VAL SER GLY ALA

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	105.11Å 105.47Å 107.94Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	35.50 – 3.02 48.01 – 3.02	Depositor EDS
% Data completeness (in resolution range)	98.6 (35.50-3.02) 98.6 (48.01-3.02)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.32 (at 3.01Å)	Xtriage
Refinement program	PHENIX (1.20.1_4487: ???)	Depositor
R, R_{free}	0.203 , 0.248 0.209 , 0.255	Depositor DCC
R_{free} test set	1163 reflections (4.89%)	wwPDB-VP
Wilson B-factor (Å^2)	76.5	Xtriage
Anisotropy	0.085	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 45.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.007 for -h,l,k 0.031 for -l,-k,-h 0.014 for k,h,-l 0.000 for k,l,h 0.000 for l,h,k	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	8011	wwPDB-VP
Average B, all atoms (Å^2)	73.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.21% 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: YJI

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	0.27	0/4020	0.52	0/5503
1	B	0.27	0/4060	0.53	0/5558
All	All	0.27	0/8080	0.52	0/11061

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	3933	0	3896	136	0
1	B	3971	0	3928	137	0
2	A	24	25	0	0	0
2	B	24	25	0	0	0
3	A	4	0	0	0	0
3	B	5	0	0	0	0
All	All	7961	50	7824	271	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 17.

The worst 5 of 271 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:2773:LEU:HD13	1:B:2832:ALA:HB1	1.51	0.90
1:B:2627:ARG:HG3	1:B:2636:VAL:HG21	1.57	0.86
1:A:2571:LEU:HD23	1:A:2576:ARG:HG2	1.56	0.86
1:A:2622:LEU:HD13	1:A:2676:LEU:HD22	1.57	0.85
1:B:2540:ALA:HB2	1:B:2549:ARG:CD	2.08	0.82

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	503/528 (95%)	484 (96%)	19 (4%)	0	100	100
1	B	511/528 (97%)	490 (96%)	20 (4%)	1 (0%)	47	81
All	All	1014/1056 (96%)	974 (96%)	39 (4%)	1 (0%)	51	85

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	2908	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	410/421 (97%)	405 (99%)	5 (1%)	71 89
1	B	413/421 (98%)	407 (98%)	6 (2%)	65 86
All	All	823/842 (98%)	812 (99%)	11 (1%)	69 88

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	2906	ARG
1	B	2908	VAL
1	B	2910	ARG
1	B	2909	SER
1	A	2913	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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)

2 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	YJI	B	3101	1	17,23,24	2.07	5 (29%)	22,29,32	1.43	3 (13%)
2	YJI	A	3101	1	17,23,24	2.11	4 (23%)	22,29,32	1.42	2 (9%)

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	YJI	B	3101	1	-	3/27/29/30	-
2	YJI	A	3101	1	-	13/27/29/30	-

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	3101	YJI	C10-N11	5.44	1.45	1.33
2	A	3101	YJI	C14-N15	5.26	1.45	1.33
2	B	3101	YJI	C14-N15	5.21	1.45	1.33
2	B	3101	YJI	C10-N11	5.12	1.44	1.33
2	B	3101	YJI	O23-C10	-2.45	1.18	1.23

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	3101	YJI	C12-C13-C14	3.24	117.76	112.36
2	B	3101	YJI	C12-C13-C14	2.94	117.25	112.36
2	A	3101	YJI	C19-S18-C17	2.50	109.46	101.87
2	B	3101	YJI	C19-S18-C17	2.04	108.05	101.87
2	B	3101	YJI	C13-C14-N15	2.03	119.84	116.42

There are no chirality outliers.

5 of 16 torsion outliers are listed below:

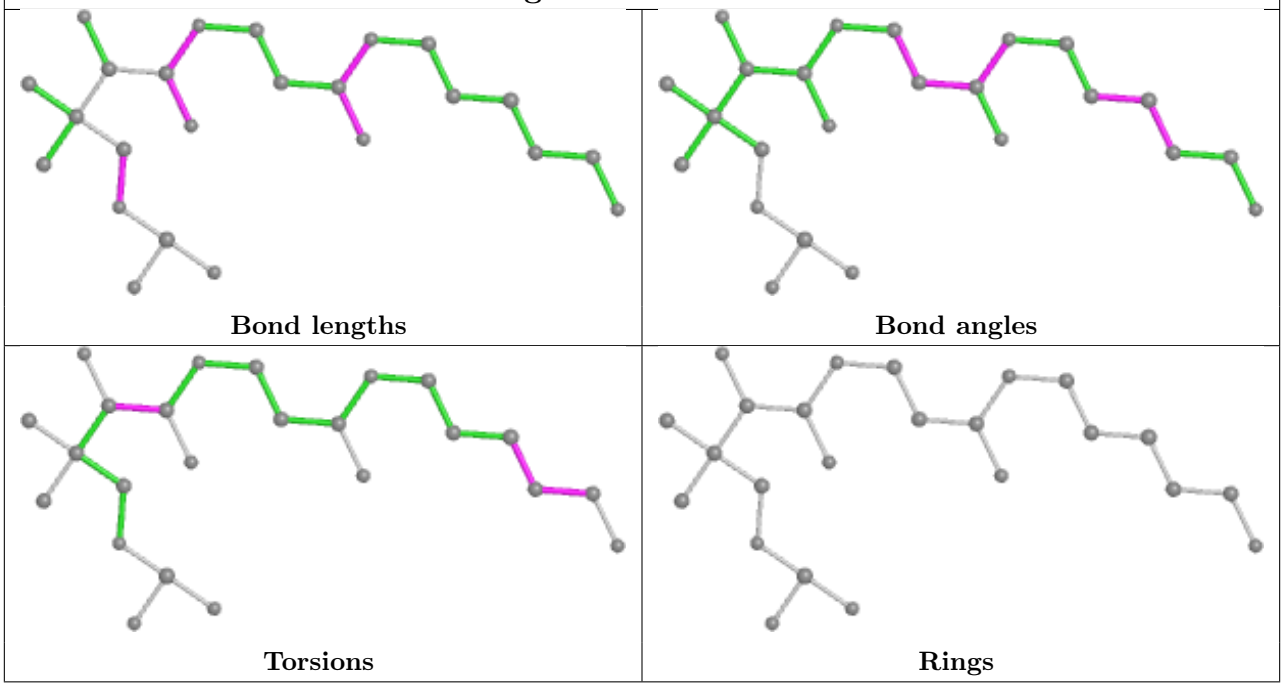
Mol	Chain	Res	Type	Atoms
2	A	3101	YJI	C03-C02-C09-C10
2	A	3101	YJI	C03-C02-C09-O24
2	A	3101	YJI	C09-C10-N11-C12
2	A	3101	YJI	N11-C12-C13-C14
2	B	3101	YJI	S18-C19-C20-O21

There are no ring outliers.

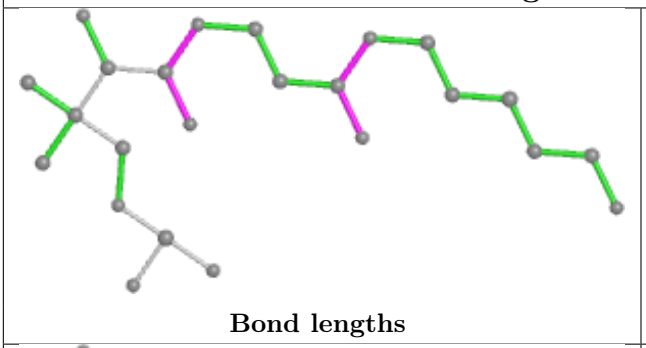
No monomer is involved in short contacts.

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.

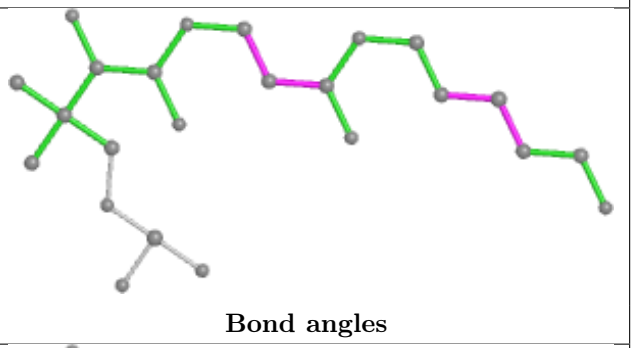
Ligand YJI B 3101



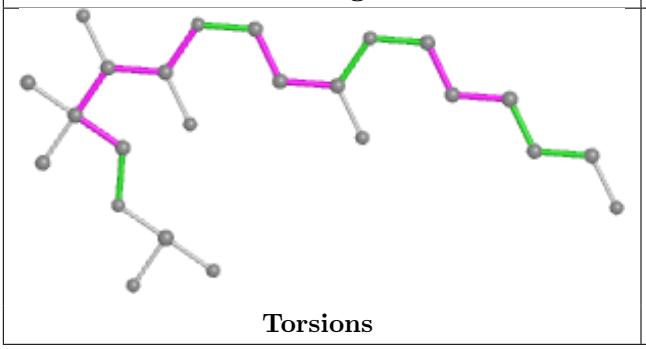
Ligand YJI A 3101



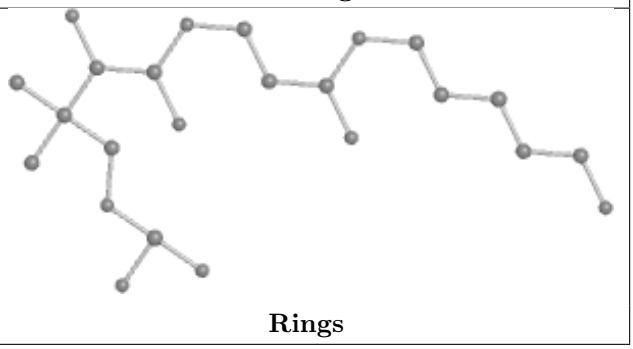
Bond lengths



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	509/528 (96%)	-0.17	1 (0%) 95 87	50, 77, 103, 126	0
1	B	515/528 (97%)	-0.27	0 100 100	44, 66, 94, 126	0
All	All	1024/1056 (96%)	-0.22	1 (0%) 95 89	44, 71, 100, 126	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	2796	LEU	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 monosaccharides 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.

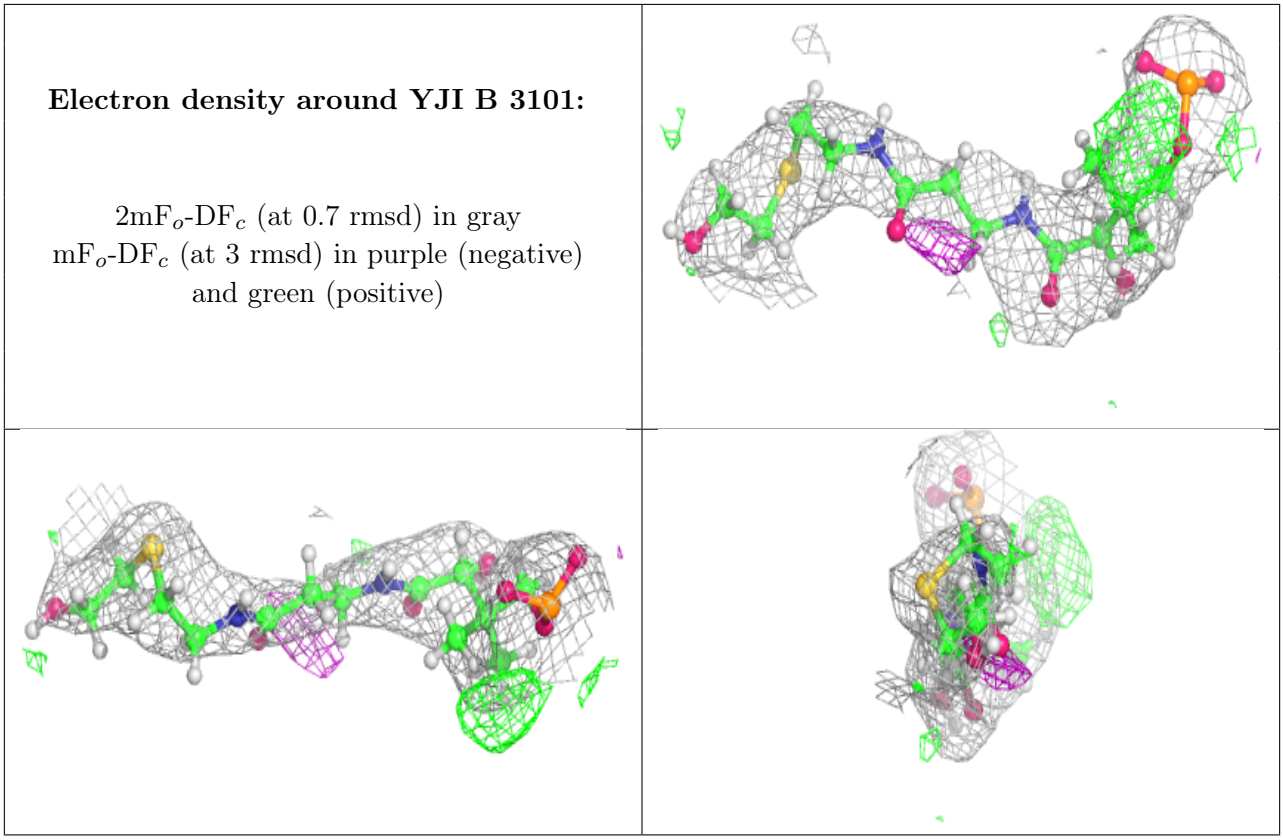
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	YJI	B	3101	24/25	0.86	0.28	80,100,117,133	0
2	YJI	A	3101	24/25	0.88	0.31	81,106,132,142	0

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.

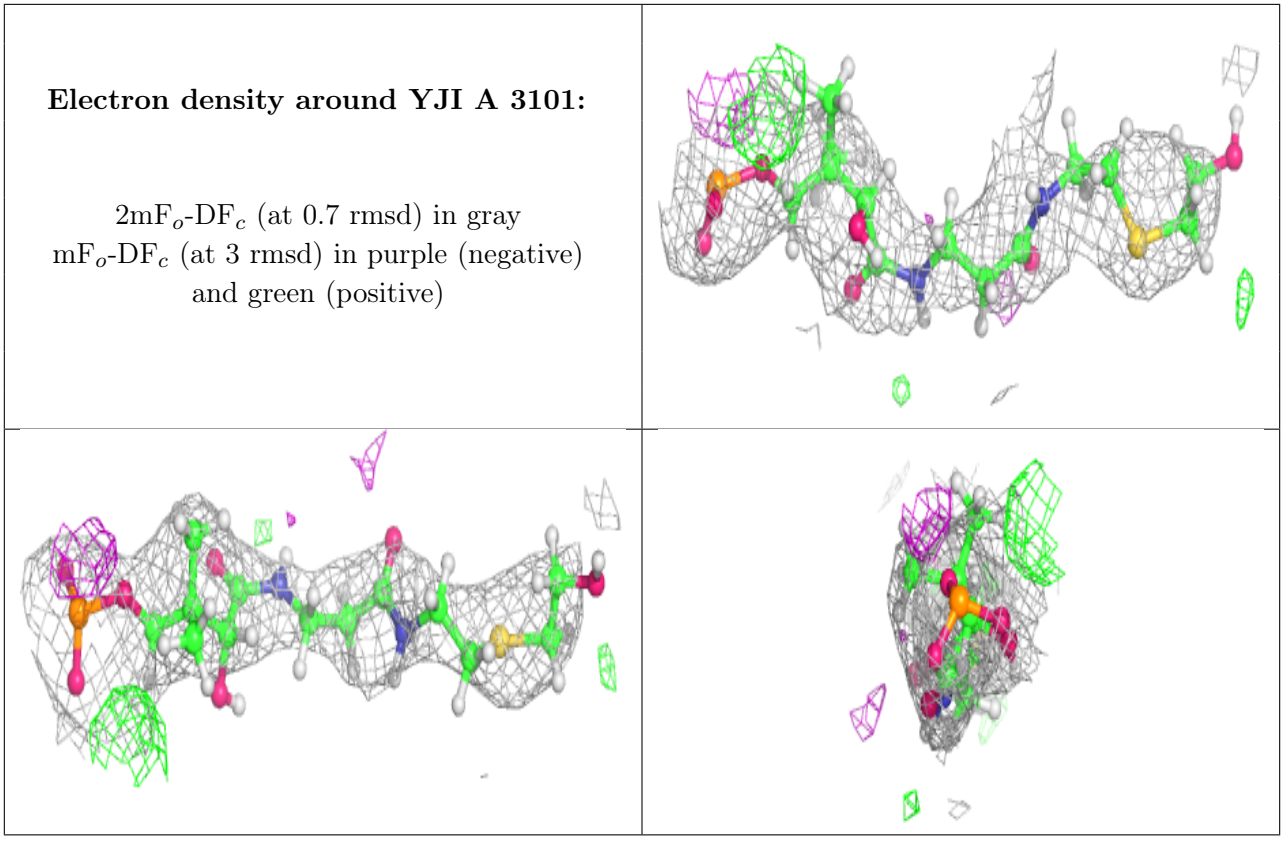
Electron density around YJI B 3101:

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



Electron density around YJI A 3101:

$2mF_o$-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.