

# wwPDB X-ray Structure Validation Summary Report (i)

#### Jun 21, 2021 – 09:11 PM BST

PDB ID : 5F9T

Title: Crystal Structure of Streptococcus pneumoniae NanC, covalent complex with

a fluorinated Neu5Ac derivative

Authors: Owen, C.D.; Lukacik, P.; Potter, J.A.; Walsh, M.; Taylor, G.L.

Deposited on : 2015-12-10

Resolution : 2.05 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 following versions of software and data (see references (1)) 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 : 2.20 report : 1.1.7 (2018)

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)

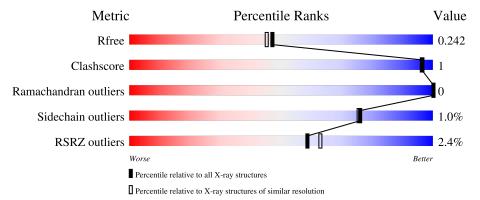
Validation Pipeline (wwPDB-VP) : 2.20

## 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 2.05 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	Similar resolution
Metric	$(\# \mathrm{Entries})$	$(\#  ext{Entries},  ext{resolution range}( ext{Å}))$
$R_{free}$	130704	1692 (2.04-2.04)
Clashscore	141614	1773 (2.04-2.04)
Ramachandran outliers	138981	1752 (2.04-2.04)
Sidechain outliers	138945	1752 (2.04-2.04)
RSRZ outliers	127900	1672 (2.04-2.04)

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	659	97%	• •
1	В	659	97%	



# 2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 11106 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.

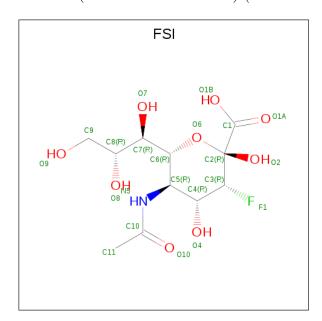
• Molecule 1 is a protein called Neuraminidase C.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
1	A	659	Total 5230	C 3321	N 881	O 1017	S 11	0	1	0
1	В	659	Total 5231	C 3321	N 881	O 1019	S 10	0	1	0

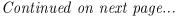
There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
Α	741	LEU	-	expression tag	UNP A0A0H2UQE4
В	741	LEU	-	expression tag	UNP A0A0H2UQE4

• Molecule 2 is 5-acetamido-3,5-dideoxy-3-fluoro-D-erythro-alpha-L-manno-non-2-ulopyranos onic acid (three-letter code: FSI) (formula: C<sub>11</sub>H<sub>18</sub>FNO<sub>9</sub>).



$\mathbf{Mol}$	Chain	Residues	${f Atoms}$			ZeroOcc	AltConf		
2	٨	1	Total	С	F	N	О	0	0
<i>Z</i>	Α	1	21	11	1	1	8	0	0

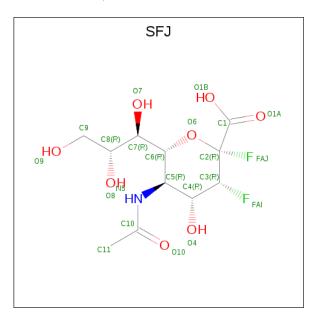




Continued from previous page...

Mol	Chain	Residues	${f Atoms}$			ZeroOcc	AltConf		
า	D	1	Total	С	F	N	О	0	0
<i>Z</i>	Ъ	1	21	11	1	1	8	U	

• Molecule 3 is (2R,3R,4R,5R,6R)-5-acetamido-2,3-difluoro-4-hydroxy-6-[(1R,2R)-1,2,3-tr ihydroxypropyl]tetrahydro-2H-pyran-2-carboxylic acid (three-letter code: SFJ) (formula:  $C_{11}H_{17}F_2NO_8$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf		
9	Α	1	Total	С	F	N	О	0	0
3	A	1	22	11	2	1	8	U	
9	D	1	Total	С	F	N	О	0	0
3	Б	1	22	11	2	1	8		0

• Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).





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

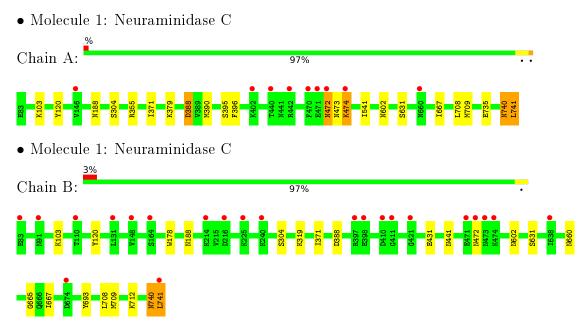
#### • Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	322	Total O 322 322	0	0
5	В	231	Total O 231 231	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.





# 4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants	100.24Å 74.84Å 113.35Å	Depositor
a, b, c, $\alpha$ , $\beta$ , $\gamma$	90.00° 96.31° 90.00°	Depositor
Resolution (Å)	30.00 - 2.05	Depositor
Resolution (A)	23.63 - 2.05	EDS
% Data completeness	97.4 (30.00-2.05)	Depositor
(in resolution range)	97.5 (23.63-2.05)	EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$< I/\sigma(I) > 1$	2.76 (at 2.06Å)	Xtriage
Refinement program	REFMAC 5.8.0103	Depositor
P. P.	0.198 , $0.235$	Depositor
$R, R_{free}$	0.206 , $0.242$	DCC
$R_{free}$ test set	5098 reflections $(5.00%)$	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	25.0	Xtriage
Anisotropy	0.431	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$ , $B_{sol}(Å^2)$	0.36 , 43.0	EDS
L-test for twinning <sup>2</sup>	$ < L > = 0.48, < L^2> = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	11106	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	33.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 4.50% of the height of the origin peak. No significant pseudotranslation is detected.

<sup>&</sup>lt;sup>2</sup>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.



<sup>&</sup>lt;sup>1</sup>Intensities estimated from amplitudes.

# 5 Model quality (i)

## 5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, FSI, SFJ

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	Bo	ond angles
MIOI	Chain	RMSZ	# Z >5	RMSZ	# Z  > 5
1	A	0.53	0/5350	0.72	$2/7252 \ (0.0\%)$
1	В	0.52	0/5351	0.72	0/7253
All	All	0.53	0/10701	0.72	2/14505~(0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1
1	В	0	1
All	All	0	2

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$\mathbf{Observed}(^o)$	$\operatorname{Ideal}({}^o)$
1	A	355	ARG	NE-CZ-NH1	5.45	123.02	120.30
1	A	388	ASP	CB-CG-OD1	5.36	123.12	118.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	Α	740	ASN	Peptide
1	В	740	ASN	Peptide



## 5.2 Too-close contacts (i)

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	Α	5230	0	5116	11	0
1	В	5231	0	5112	9	0
2	A	21	0	16	0	0
2	В	21	0	16	0	0
3	A	22	0	0	0	0
3	В	22	0	0	0	0
4	В	6	0	8	0	0
5	A	322	0	0	0	0
5	В	231	0	0	0	0
All	All	11106	0	10268	20	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 1.

The worst 5 of 20 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic	Clash
Atom-1	Atom-2	${f distance}\;({f \AA})$	$ \text{overlap } (\text{\AA})$
1:B:431[B]:GLU:HG2	1:B:431[B]:GLU:O	1.86	0.74
1:A:474:LYS:HA	1:A:474:LYS:HE3	1.89	0.53
1:A:304:SER:HB2	1:A:709:MET:HE1	1.93	0.50
1:B:667:ILE:HD11	1:B:708:LEU:HD22	1.93	0.50
1:A:740:ASN:O	1:A:741:LEU:HB2	2.12	0.49

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	$658/659 \; (100\%)$	640 (97%)	18 (3%)	0	100 100		
1	В	$658/659 \; (100\%)$	642 (98%)	16 (2%)	0	100 100		
All	All	1316/1318 (100%)	1282 (97%)	34 (3%)	0	100 100		

There are no Ramachandran outliers to report.

#### 5.3.2Protein 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	574/573 (100%)	568 (99%)	6 (1%)	76	75	
1	В	574/573 (100%)	568 (99%)	6 (1%)	76	75	
All	All	1148/1146 (100%)	1136 (99%)	12 (1%)	76	75	

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	В	319	LYS
1	В	441	ASN
1	В	741	LEU
1	В	472	ASN
1	A	474	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
1	В	511	HIS
1	В	676	ASN
1	A	459	ASN
1	В	246	ASN
1	В	441	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)

5 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		
10101	Type	Chain	nes	Lilik	Counts	RMSZ	$\mid \# Z  > 2$	Counts	RMSZ	# Z  > 2
3	SFJ	A	802	-	16,22,22	0.46	0	19,33,33	0.82	0
4	GOL	В	801	-	5,5,5	0.26	0	5,5,5	0.43	0
2	FSI	A	801	1	18,21,22	0.84	1 (5%)	18,30,33	0.98	1 (5%)
3	SFJ	В	803	-	16,22,22	0.44	0	19,33,33	0.97	1 (5%)
2	FSI	В	802	1	18,21,22	0.63	0	18,30,33	0.84	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
3	SFJ	A	802	-	-	2/14/43/43	0/1/1/1
4	GOL	В	801	-	-	2/4/4/4	-
2	FSI	A	801	1	-	4/14/38/43	0/1/1/1
3	SFJ	В	803	_	-	1/14/43/43	0/1/1/1

Continued on next page...



Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	FSI	В	802	1	-	2/14/38/43	0/1/1/1

#### All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$\operatorname{Observed}(\operatorname{\AA})$	$\mathbf{Ideal}(\mathbf{\AA})$
2	A	801	FSI	C3-C2	3.01	1.55	1.52

#### All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$\mathbf{Observed}(^o)$	$\operatorname{Ideal}({}^o)$
2	A	801	FSI	F1-C3-C2	3.06	111.96	108.26
3	В	803	SFJ	O6-C6-C5	2.48	112.20	109.78

There are no chirality outliers.

5 of 11 torsion outliers are listed below:

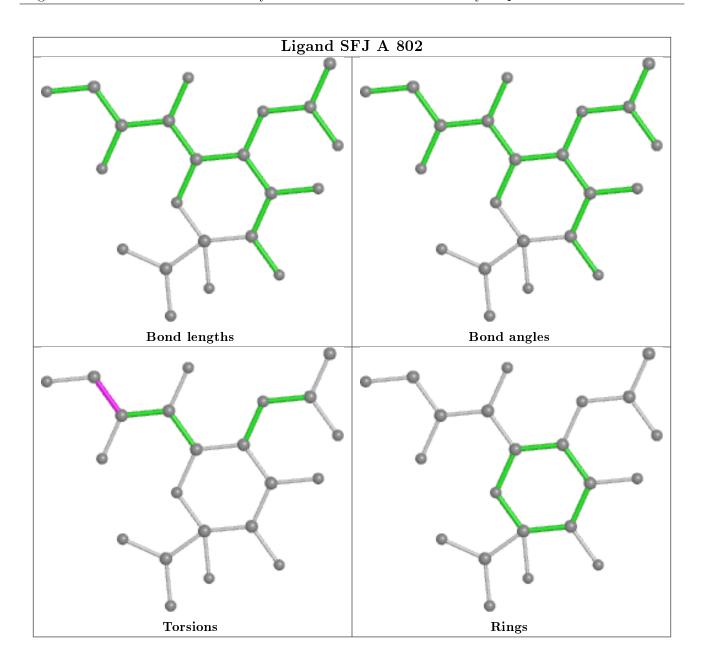
Mol	Chain	Res	Type	Atoms
2	A	801	FSI	O7-C7-C8-C9
4	В	801	GOL	O1-C1-C2-C3
2	A	801	FSI	C6-C7-C8-C9
2	A	801	FSI	C6-C7-C8-O8
3	A	802	SFJ	O8-C8-C9-O9

There are no ring outliers.

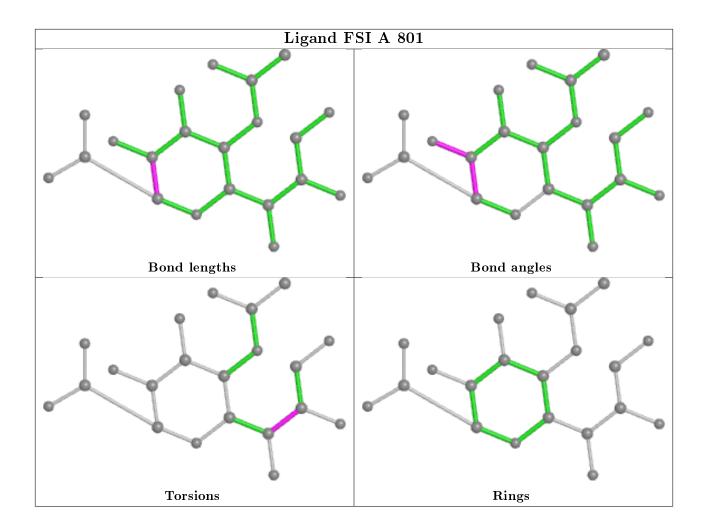
No monomer is involved in short contacts.

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.

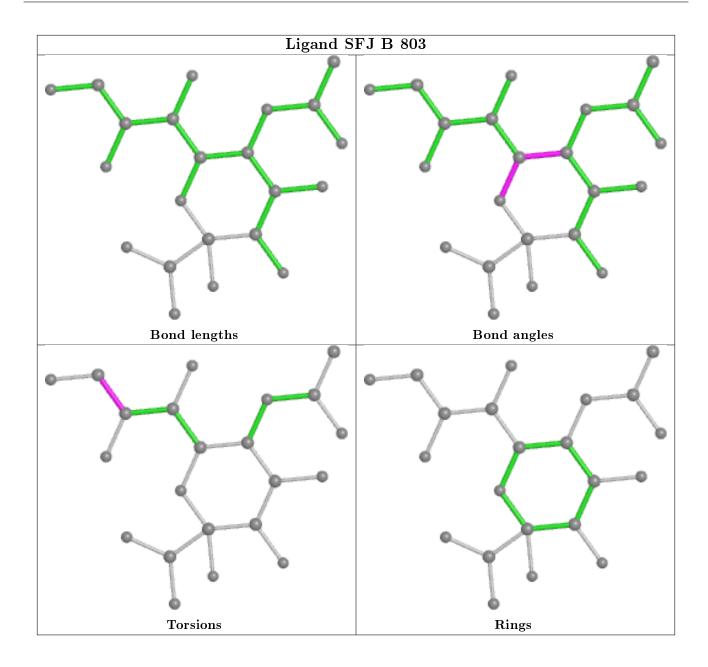




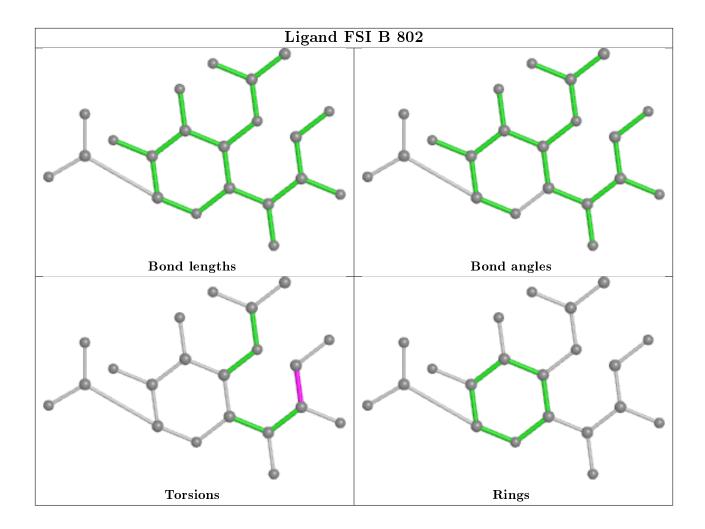












# 5.7 Other polymers (i)

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,  $95^{th}$  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 $>$	$\#\mathrm{RSRZ}{>}2$	$OWAB(A^2)$	Q < 0.9
1	A	$659/659 \; (100\%)$	-0.19	9 (1%) 75 78	16, 27, 51, 116	0
1	В	$659/659 \; (100\%)$	0.03	22 (3%) 46 50	18, 34, 58, 87	0
All	All	1318/1318 (100%)	-0.08	31 (2%) 59 63	16, 31, 56, 116	0

The worst 5 of 31 RSRZ outliers are listed below:

Mol	Chain	${f Res}$	Type	RSRZ
1	A	472	ASN	6.3
1	A	471	GLU	5.7
1	A	442	ARG	4.6
1	В	741	LEU	4.1
1	В	472	ASN	4.0

## 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 monosaccharides in this entry.

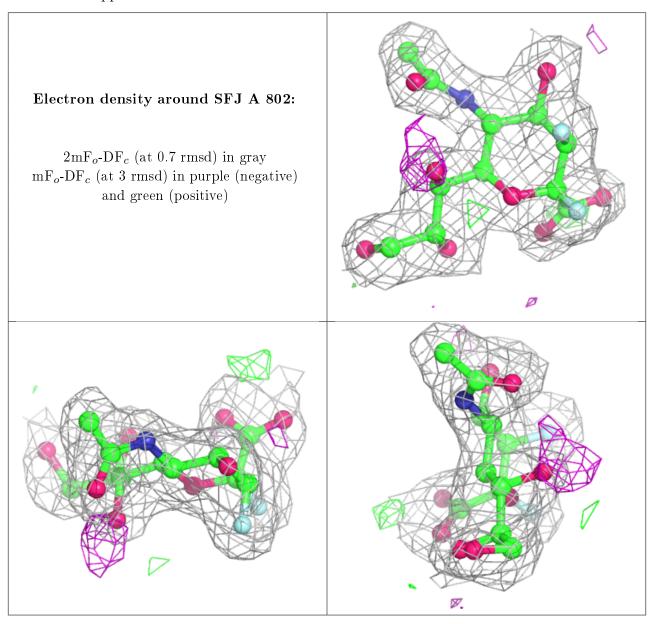
## 6.4 Ligands (i)

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,  $95^{th}$  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.

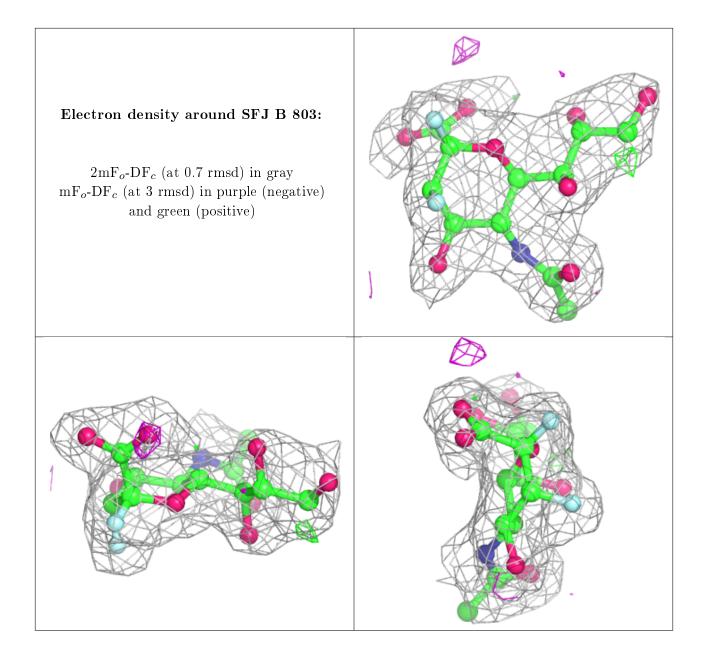


Mol	Type	Chain	Res	Atoms	RSCC	RSR	$\mathbf{B} ext{-}\mathbf{factors}(\mathbf{\mathring{A}}^2)$	Q < 0.9
3	SFJ	A	802	22/22	0.91	0.13	22,31,39,53	0
3	SFJ	В	803	22/22	0.92	0.18	26,37,54,64	0
4	GOL	В	801	6/6	0.92	0.13	33,36,43,52	0
2	FSI	A	801	21/22	0.95	0.10	15,20,32,59	0
2	FSI	В	802	21/22	0.96	0.09	15,23,37,53	0

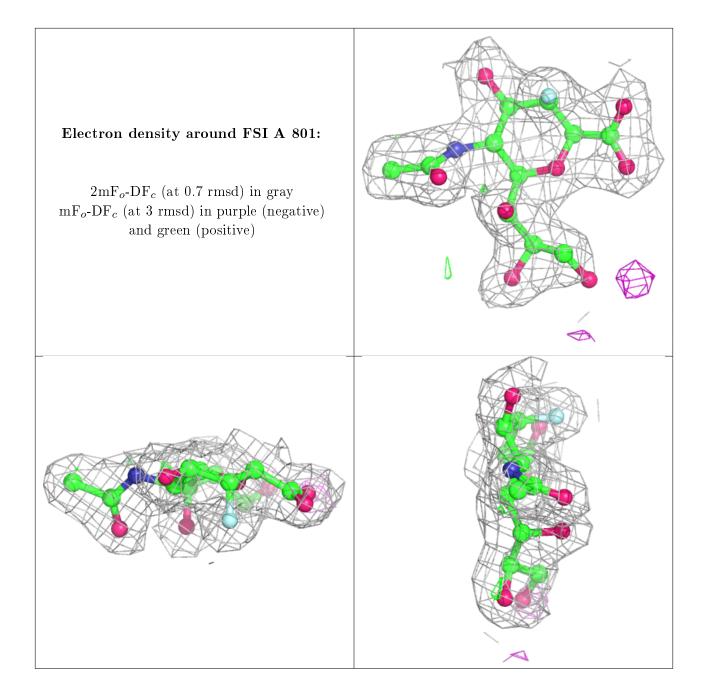
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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



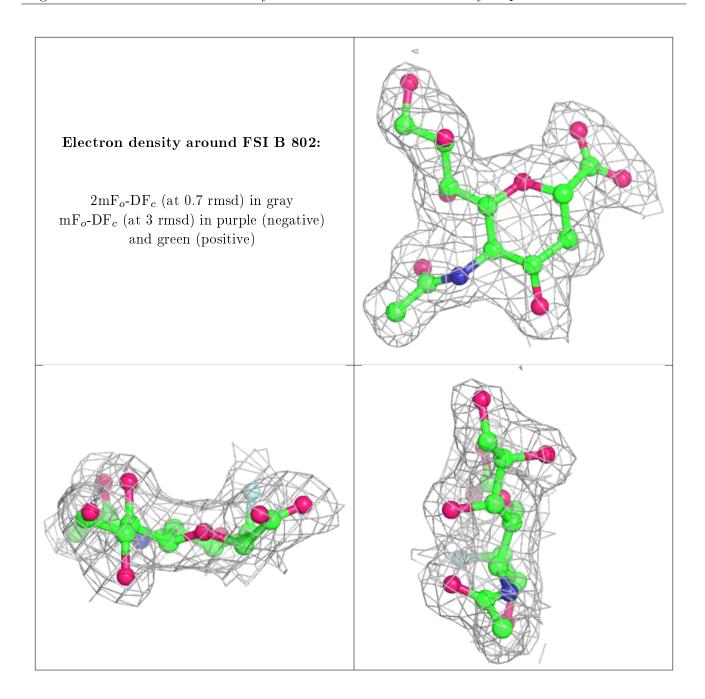












# 6.5 Other polymers (i)

There are no such residues in this entry.

