WORLDWIDE PDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

May 17, 2020 – 03:53 am BST

PDB ID	:	6EN8
Title	:	SaFadR in complex with dsDNA
Authors	:	Valegard, K.
Deposited on	:	2017-10-04
Resolution	:	3.29 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

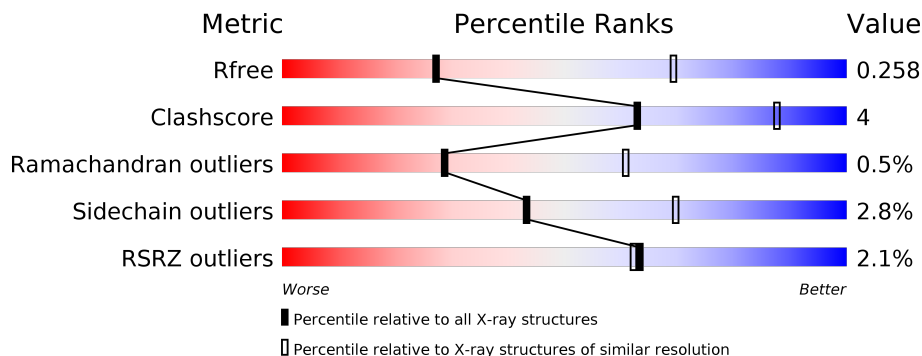|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 3.29 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	Percentile Ranks	Value
Rfree		0.258
Clashscore		4
Ramachandran outliers		0.5%
Sidechain outliers		2.8%
RSRZ outliers		2.1%

Worse   *Better*

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1149 (3.34-3.26)
Clashscore	141614	1205 (3.34-3.26)
Ramachandran outliers	138981	1183 (3.34-3.26)
Sidechain outliers	138945	1182 (3.34-3.26)
RSRZ outliers	127900	1115 (3.34-3.26)

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 on 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	196	85% 13% •
1	B	196	% 73% 18% • 6%
1	C	196	2% 86% 11% • •
1	D	196	3% 76% 14% 11%
1	E	196	4% 83% 14% • •
1	F	196	4% 89% 7% • •

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	Q	22	73% 23% 5%
2	Y	22	73% 23% 5%
3	X	22	55% 41% 5%
3	Z	22	64% 32% 5%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 11131 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 Transcriptional regulator TetR family.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	191	Total 1591	C 1028	N 269	O 287	Se 7	0	0	0
1	B	184	Total 1541	C 997	N 260	O 277	Se 7	0	0	0
1	C	191	Total 1591	C 1028	N 269	O 287	Se 7	0	0	0
1	D	175	Total 1457	C 947	N 242	O 261	Se 7	0	0	0
1	E	191	Total 1592	C 1028	N 272	O 285	Se 7	0	0	0
1	F	189	Total 1573	C 1017	N 264	O 285	Se 7	0	0	0

- Molecule 2 is a DNA chain called DNA (5'-D(*GP*TP*CP*GP*AP*CP*TP*CP*AP*AP* AP*AP*AP*TP*CP*AP*AP*GP*TP*AP*G)-3').
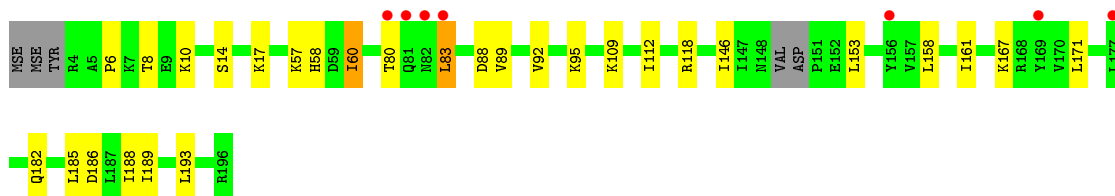
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	Q	21	Total 430	C 206	N 85	O 119	P 20	0	0	0
2	Y	21	Total 430	C 206	N 85	O 119	P 20	0	0	0

- Molecule 3 is a DNA chain called DNA (5'-D(*CP*TP*AP*CP*TP*TP*GP*AP*TP*TP* TP*TP*TP*GP*AP*GP*TP*CP*GP*AP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	X	21	Total 425	C 206	N 70	O 129	P 20	0	0	0
3	Z	21	Total 428	C 206	N 70	O 131	P 21	0	0	0

- Molecule 4 is UNKNOWN LIGAND (three-letter code: UNL) (formula:).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	F	1	Total	C	N	O	P	0	0
			33	12	5	13	3		
4	E	1	Total	C	N	O	P	0	0
			40	16	6	15	3		

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA and DNA 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: Transcriptional regulator TetR family

Chain A:

• Molecule 1: Transcriptional regulator TetR family

Chain B:

• Molecule 1: Transcriptional regulator TetR family

Chain C:

• Molecule 1: Transcriptional regulator TetR family

Chain D:

• Molecule 1: Transcriptional regulator TetR family

Chain E:

4%

83%   14%

MSE MSE TYR R4 A5 P6 K7 T8 E9 K10 S14 K17 K57 H58 D59 I60 T80 Q81 N82 L83 D88 V89 V92 K95 K109 I112 R118 I146 I147 N148 VAL ASP P151 E152 L153 Y156 V157 L158 I161 K167 R168 Y169 V170 L171 L177

Q182 L185 D186 L187 I188 I189 L193 R196

• Molecule 1: Transcriptional regulator TetR family

Chain F:

4%

89%   7%

MSE MSE TYR ARG ALA P6 I18 L49 K57 I60 N82 R86 Y107 Y108 K109 I112 Y133 L137 P151 L154 V157 L177 T178 L179 D184 L185 D186 L187 I188 I189 T194 PRO ARG

• Molecule 2: DNA (5'-D(*GP*TP*CP*GP*AP*CP*TP*CP*AP*AP*AP*AP*AP*TP*CP*AP *AP*GP*TP*AP*G)-3')

Chain Q:

73%   23%   5%

UNK G1 T2 C3 G4 A5 C8 A9 A12 G21

• Molecule 2: DNA (5'-D(*GP*TP*CP*GP*AP*CP*TP*CP*AP*AP*AP*AP*AP*TP*CP*AP *AP*GP*TP*AP*G)-3')

Chain Y:

73%   23%   5%

UNK G1 A5 A11 A12 A13 T14 G21

• Molecule 3: DNA (5'-D(*CP*TP*AP*CP*TP*TP*GP*AP*TP*TP*TP*TP*TP*GP*AP*GP *TP*CP*GP*AP*C)-3')

Chain X:

55%   41%   5%

UNK C1 G7 A8 T9 T10 T11 T12 T13 G14 A15 A20 C21

• Molecule 3: DNA (5'-D(*CP*TP*AP*CP*TP*TP*GP*AP*TP*TP*TP*TP*TP*GP*AP*GP *TP*CP*GP*AP*C)-3')

Chain Z:

64%   32%   5%

UNK C1 T2 A3 T12 T13 T17 C18 G19 A20 C21

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	54.85Å 178.31Å 266.70Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.41 – 3.29 49.41 – 3.29	Depositor EDS
% Data completeness (in resolution range)	99.5 (49.41-3.29) 99.6 (49.41-3.29)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	2.48 (at 3.33Å)	Xtriage
Refinement program	BUSTER 2.10.1	Depositor
R, R_{free}	0.202 , 0.238 0.222 , 0.258	Depositor DCC
R_{free} test set	2035 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	97.4	Xtriage
Anisotropy	0.857	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 100.5	EDS
L-test for twinning[2]	$<|L|> = 0.46$, $<L^2> = 0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	11131	wwPDB-VP
Average B, all atoms (Å^2)	139.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.66% 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: UNL

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	0.38	0/1617	0.57	0/2161
1	B	0.40	0/1565	0.60	0/2091
1	C	0.40	0/1617	0.55	0/2161
1	D	0.40	0/1480	0.56	0/1978
1	E	0.41	0/1617	0.58	0/2158
1	F	0.40	0/1598	0.56	0/2135
2	Q	0.99	0/484	0.97	0/745
2	Y	0.90	0/484	0.98	0/745
3	X	1.01	0/474	1.07	1/730 (0.1%)
3	Z	1.10	0/477	1.05	0/734
All	All	0.55	0/11413	0.68	1/15638 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	X	13	DT	O4'-C4'-C3'	-6.61	101.86	104.50

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	1591	0	1621	13	0
1	B	1541	0	1565	21	0
1	C	1591	0	1621	14	0
1	D	1457	0	1481	11	0
1	E	1592	0	1625	20	0
1	F	1573	0	1601	9	0
2	Q	430	0	237	5	0
2	Y	430	0	237	4	0
3	X	425	0	242	6	0
3	Z	428	0	241	5	0
4	E	40	0	0	0	0
4	F	33	0	0	0	0
All	All	11131	0	10471	93	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.

The worst 5 of 93 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)
2:Y:11:DA:H2'	2:Y:12:DA:C8	2.08	0.87
3:Z:17:DT:H2"	3:Z:18:DC:O5'	1.88	0.74
1:B:101:MSE:HE3	1:B:166:GLY:HA3	1.72	0.71
3:X:14:DG:H2"	3:X:15:DA:C8	2.26	0.69
3:X:10:DT:H2"	3:X:11:DT:O5'	1.95	0.67

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	189/196 (96%)	183 (97%)	6 (3%)	0	100	100

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	180/196 (92%)	170 (94%)	9 (5%)	1 (1%)	25	57
1	C	189/196 (96%)	180 (95%)	9 (5%)	0	100	100
1	D	171/196 (87%)	160 (94%)	10 (6%)	1 (1%)	25	57
1	E	187/196 (95%)	176 (94%)	9 (5%)	2 (1%)	14	45
1	F	187/196 (95%)	177 (95%)	9 (5%)	1 (0%)	29	61
All	All	1103/1176 (94%)	1046 (95%)	52 (5%)	5 (0%)	29	61

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	F	82	ASN
1	D	175	SER
1	E	83	LEU
1	B	175	SER
1	E	6	PRO

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	174/169 (103%)	170 (98%)	4 (2%)	50	73
1	B	169/169 (100%)	159 (94%)	10 (6%)	19	49
1	C	174/169 (103%)	172 (99%)	2 (1%)	73	85
1	D	159/169 (94%)	151 (95%)	8 (5%)	24	55
1	E	173/169 (102%)	172 (99%)	1 (1%)	86	91
1	F	172/169 (102%)	168 (98%)	4 (2%)	50	73
All	All	1021/1014 (101%)	992 (97%)	29 (3%)	43	70

5 of 29 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	180	LYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	D	45	VAL
1	F	60	ILE
1	C	8	THR
1	D	87	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	67	GLN
1	C	69	ASN
1	F	69	ASN
1	B	191	ASN
1	E	191	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 carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 2 ligands modelled in this entry, 2 are unknown - leaving 0 for Mogul analysis.

There are no bond length outliers.
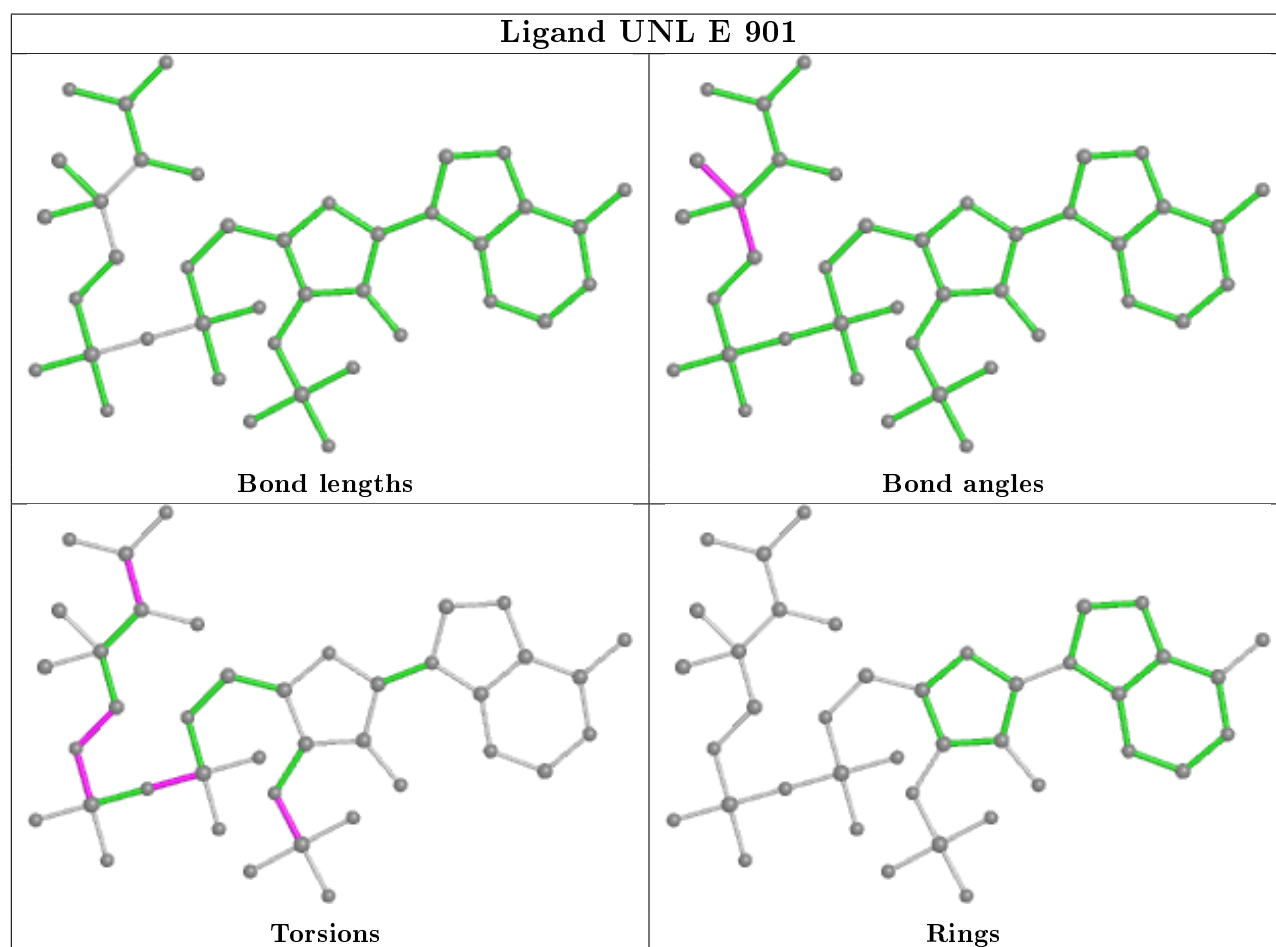
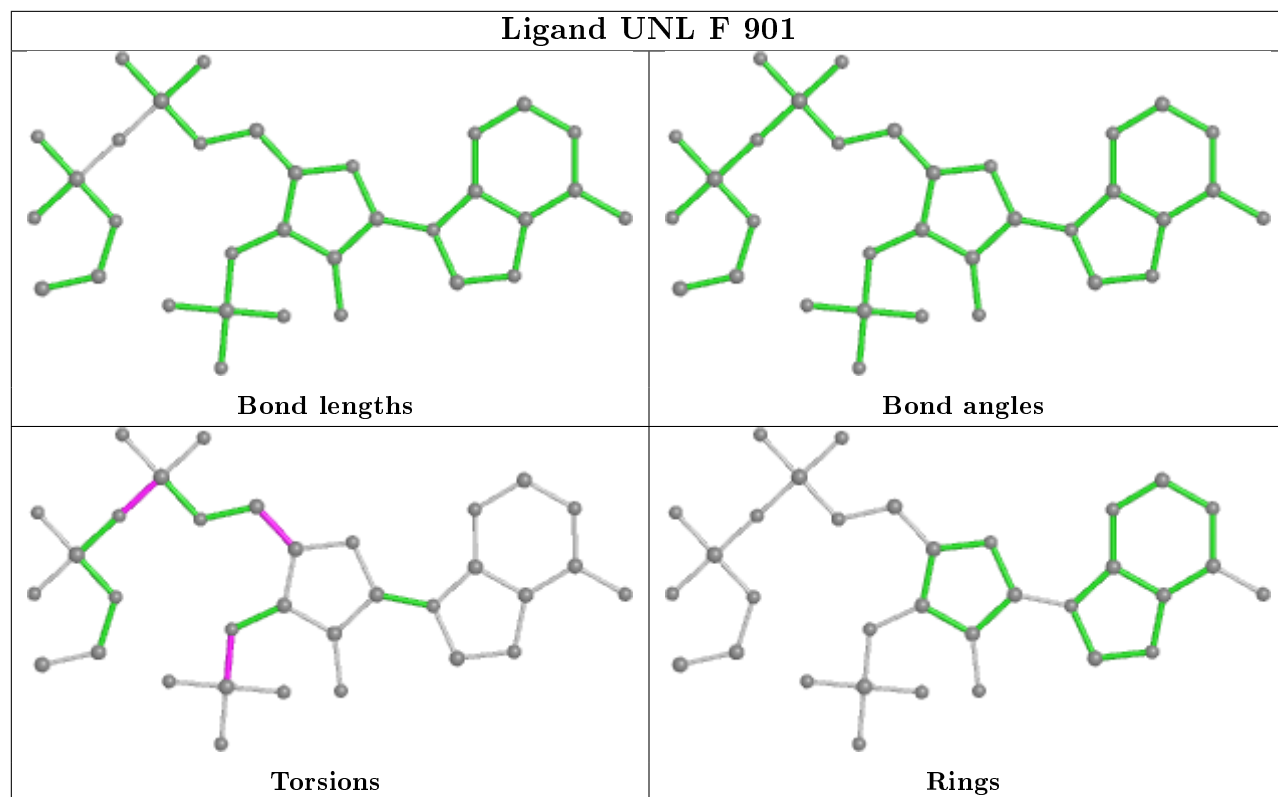There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

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 UNL E 901

Bond lengths

Bond angles

Torsions

Rings

Ligand UNL F 901

Bond lengths

Bond angles

Torsions

Rings

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(Å²)	Q<0.9
1	A	184/196 (93%)	0.10	0	100	100	85, 108, 142, 161	0
1	B	177/196 (90%)	0.20	2 (1%)	80	81	78, 106, 165, 178	0
1	C	184/196 (93%)	0.22	3 (1%)	72	70	131, 162, 191, 214	0
1	D	168/196 (85%)	0.29	6 (3%)	42	40	109, 152, 184, 201	0
1	E	184/196 (93%)	0.28	7 (3%)	40	37	92, 163, 203, 212	0
1	F	182/196 (92%)	0.38	7 (3%)	40	37	114, 166, 198, 206	0
2	Q	21/22 (95%)	-0.59	0	100	100	100, 115, 145, 150	0
2	Y	21/22 (95%)	-0.40	0	100	100	96, 108, 135, 137	0
3	X	21/22 (95%)	-0.34	0	100	100	88, 115, 132, 134	0
3	Z	21/22 (95%)	-0.52	0	100	100	100, 124, 137, 150	0
All	All	1163/1264 (92%)	0.19	25 (2%)	63	62	78, 137, 195, 214	0

The worst 5 of 25 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	81	GLN	4.4
1	D	145	GLU	4.3
1	D	143	ARG	3.7
1	D	78	THR	3.4
1	E	83	LEU	3.1

6.2 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.
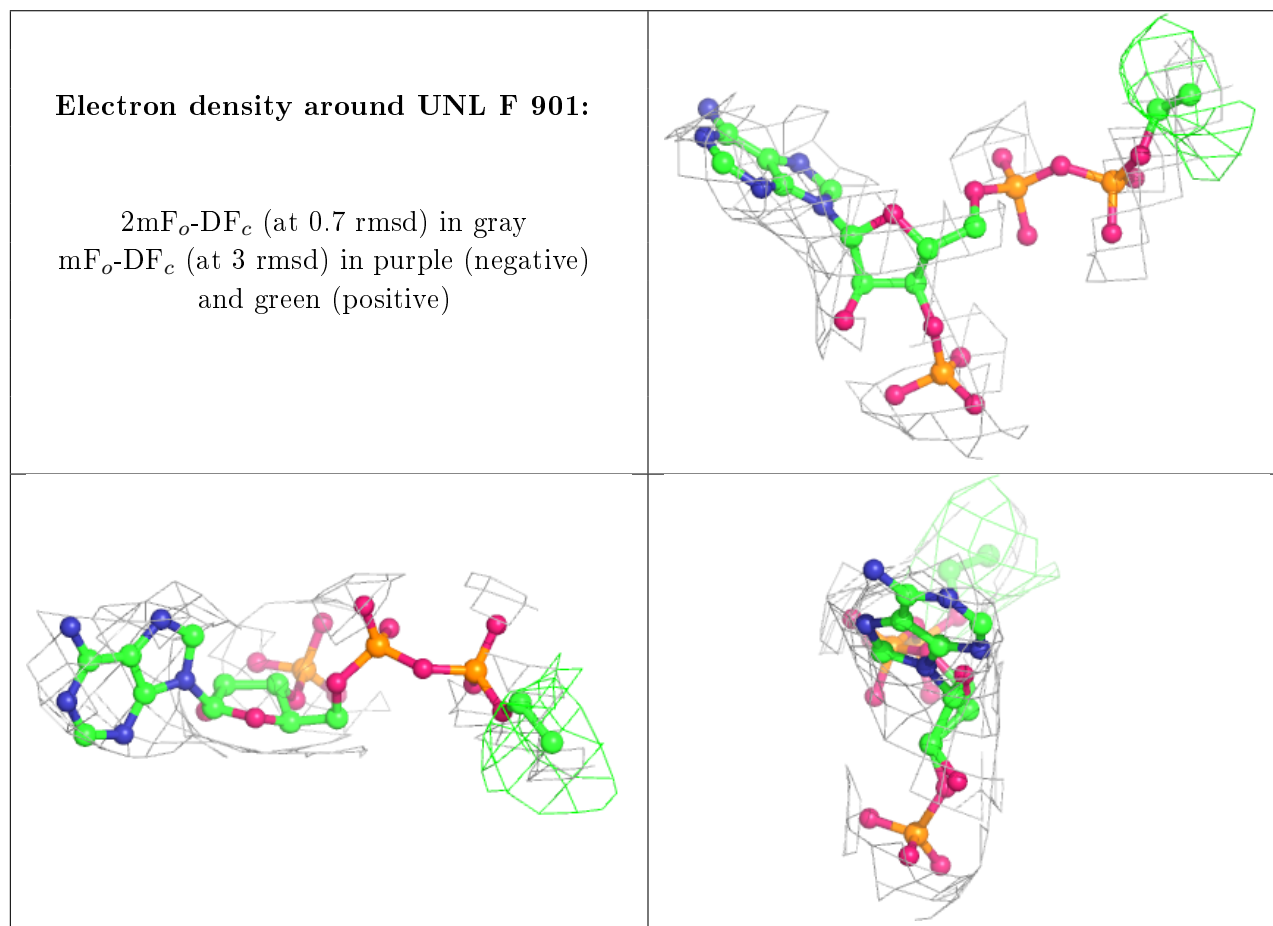
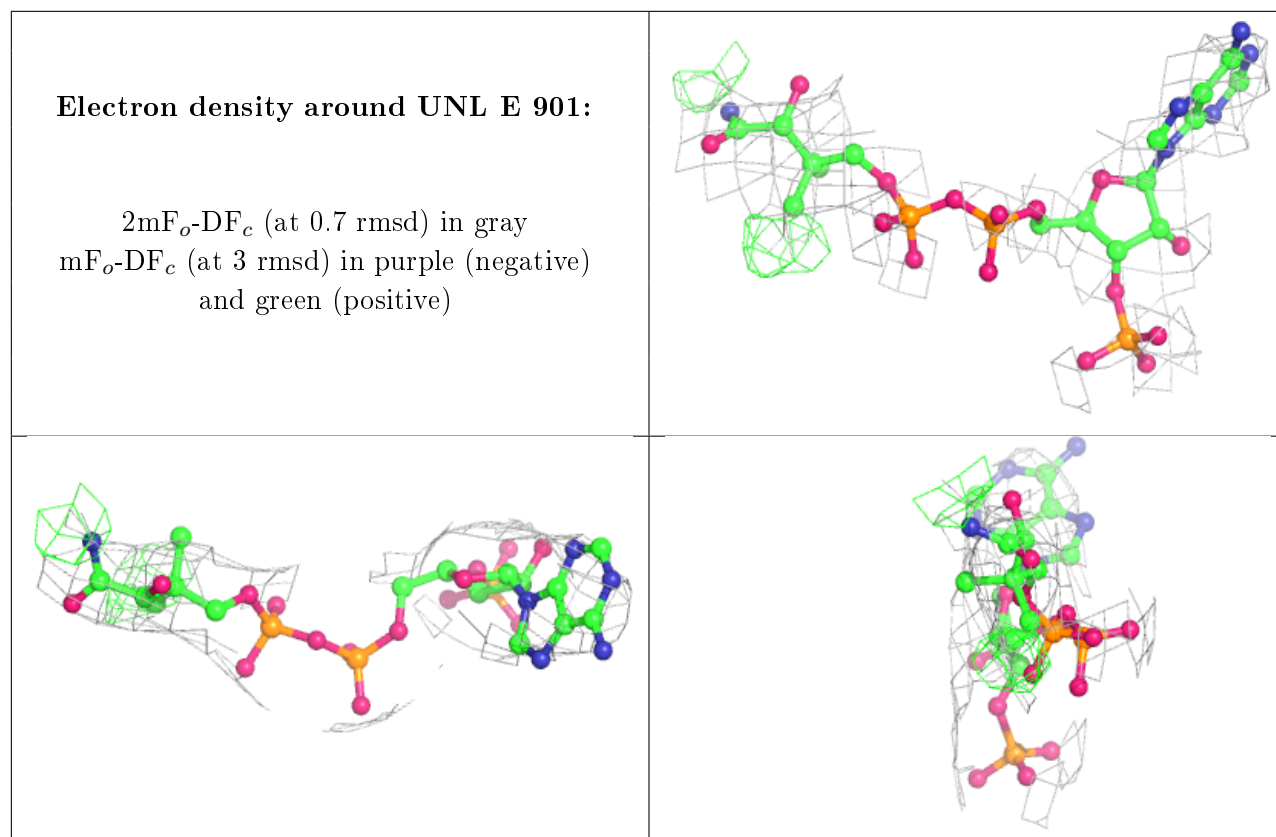## 6.3 Carbohydrates (i)

There are no carbohydrates 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($Å^2$)	Q<0.9
4	UNL	F	901	33/-	0.76	0.18	199,200,204,204	0
4	UNL	E	901	40/-	0.86	0.17	211,212,214,214	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 UNL F 901:

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

Electron density around UNL E 901:

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

6.5 Other polymers (i)

There are no such residues in this entry.