

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

#### Oct 10, 2023 – 09:05 AM EDT

PDB ID	:	7THH
Title	:	SUD-C and Ubl2 domains of SARS CoV-2 Nsp3 protein
Authors	:	Osipiuk, J.; Jedrzejczak, R.; Endres, M.; Wydorski, P.; Joachimiak, L.;
		Joachimiak, A.; Center for Structural Genomics of Infectious Diseases (CS-
		GID)
Deposited on	:	2022-01-11
Resolution	:	1.32  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:

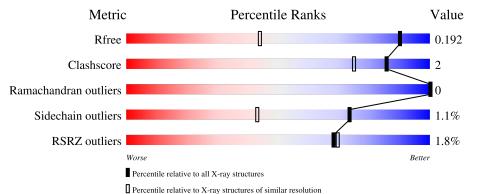
M - 1D		4 001 407
MolProbity	:	4.020-407
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

# 1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:  $X\text{-}RAY \, DIFFRACTION$ 

The reported resolution of this entry is 1.32 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	$\begin{array}{c} \textbf{Whole archive} \\ (\#\textbf{Entries}) \end{array}$	${f Similar\ resolution}\ (\#{ m Entries,\ resolution\ range}({ m \AA}))$
$R_{free}$	130704	1611(1.34-1.30)
Clashscore	141614	1667 (1.34-1.30)
Ramachandran outliers	138981	1615 (1.34-1.30)
Sidechain outliers	138945	1615 (1.34-1.30)
RSRZ outliers	127900	1580 (1.34-1.30)

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	А	131	% 92%	7%	
1	В	131	% 92%	7%	
1	С	131	% 98%	•	
1	D	131	2% 99%		•



Mol	Chain	Length	Quality of chain	
1	Е	131	<mark>6%</mark> 93%	5% ••
1	F	131	% 95%	5%•



# 2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 7374 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.

Mol	Chain	Residues		At	oms			ZeroOcc	AltConf	Trace
1	А	131	Total	С	Ν	0	$\mathbf{S}$	0	6	0
	A	101	1085	694	176	213	2	0	0	0
1	В	130	Total	С	Ν	Ο	S	0	4	0
	D	150	1062	678	174	208	2	0	4	0
1	С	131	Total	С	Ν	Ο	S	0	5	0
	U	131	1075	685	173	215	2	0	0	0
1	D	131	Total	С	Ν	Ο	$\mathbf{S}$	0	7	0
	D	101	1087	694	179	212	2	0	1	0
1	Е	130	Total	С	Ν	Ο	$\mathbf{S}$	0	1	0
L	Ľ	150	1045	667	170	206	2	0	1	0
1	F	130	Total	С	Ν	Ο	S	0	6	0
	Ľ	130	1066	684	171	209	2	0	0	U

• Molecule 1 is a protein called Papain-like protease nsp3.

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	675	SER	-	expression tag	UNP P0DTD1
А	676	ASN	-	expression tag	UNP P0DTD1
А	677	ALA	-	expression tag	UNP P0DTD1
В	675	SER	-	expression tag	UNP P0DTD1
В	676	ASN	-	expression tag	UNP P0DTD1
В	677	ALA	-	expression tag	UNP P0DTD1
С	675	SER	-	expression tag	UNP P0DTD1
С	676	ASN	-	expression tag	UNP P0DTD1
С	677	ALA	-	expression tag	UNP P0DTD1
D	675	SER	-	expression tag	UNP P0DTD1
D	676	ASN	-	expression tag	UNP P0DTD1
D	677	ALA	-	expression tag	UNP P0DTD1
Е	675	SER	-	expression tag	UNP P0DTD1
Е	676	ASN	-	expression tag	UNP P0DTD1
Е	677	ALA	-	expression tag	UNP P0DTD1
F	675	SER	-	expression tag	UNP P0DTD1
F	676	ASN	-	expression tag	UNP P0DTD1
F	676	ASN	-	- 0	UNP PODTI

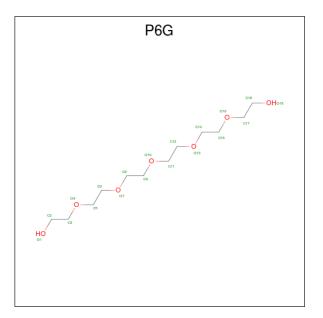


Chain	Residue	Modelled	Actual	Comment	Reference
F	677	ALA	-	expression tag	UNP P0DTD1

• Molecule 2 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	А	2	Total Cl 2 2	0	0
2	Ε	2	Total Cl 2 2	0	0
2	F	2	Total Cl 2 2	0	0

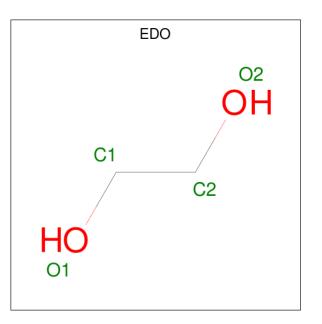
• Molecule 3 is HEXAETHYLENE GLYCOL (three-letter code: P6G) (formula:  $C_{12}H_{26}O_7$ ).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	А	1	Total         C         O           19         12         7	0	0
3	В	1	Total         C         O           19         12         7	0	0
3	С	1	Total         C         O           19         12         7	0	0

• Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula:  $C_2H_6O_2$ ).





Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	А	1	$\begin{array}{ccc} \text{Total} & \text{C} & \text{O} \\ 4 & 2 & 2 \end{array}$	0	0
4	Е	1	$\begin{array}{ccc} \text{Total} & \text{C} & \text{O} \\ 4 & 2 & 2 \end{array}$	0	0

• Molecule 5 is IODIDE ION (three-letter code: IOD) (formula: I).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	В	1	Total I 1 1	0	0
5	С	2	Total I 2 2	0	0

• Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	А	182	Total O 182 182	0	0
6	В	170	Total         O           170         170	0	0
6	С	186	Total O 186 186	0	1
6	D	109	Total O 109 109	0	3
6	Е	122	Total         O           122         122	0	0



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	F	111	Total O 111 111	0	1

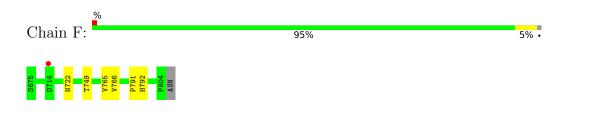


# 3 Residue-property plots (i)

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.

- Chain A: 92% 7% • • Molecule 1: Papain-like protease nsp3 Chain B: 92% 7% • • Molecule 1: Papain-like protease nsp3 Chain C: 98% • Molecule 1: Papain-like protease nsp3 Chain D: 99% • Molecule 1: Papain-like protease nsp3 Chain E: 5% •• 93% • Molecule 1: Papain-like protease nsp3
- Molecule 1: Papain-like protease nsp3







## 4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants	30.17Å 113.69Å 109.69Å	Depositor
a, b, c, $\alpha$ , $\beta$ , $\gamma$	$90.00^{\circ}$ $93.02^{\circ}$ $90.00^{\circ}$	Depositor
Resolution (Å)	49.39 - 1.32	Depositor
Resolution (A)	49.34 - 1.32	EDS
% Data completeness	98.6 (49.39-1.32)	Depositor
(in resolution range)	98.6(49.34-1.32)	EDS
R <sub>merge</sub>	0.14	Depositor
R <sub>sym</sub>	(Not available)	Depositor
$< I/\sigma(I) > 1$	$2.82 (at 1.32 \text{\AA})$	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
D D.	0.145 , $0.191$	Depositor
$R, R_{free}$	0.146 , $0.192$	DCC
$R_{free}$ test set	8728 reflections $(5.14%)$	wwPDB-VP
Wilson B-factor $(Å^2)$	17.0	Xtriage
Anisotropy	0.444	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$ , $B_{sol}(Å^2)$	0.34 , $42.3$	EDS
L-test for twinning <sup>2</sup>	$ \langle L  \rangle = 0.49, \langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.98	EDS
Total number of atoms	7374	wwPDB-VP
Average B, all atoms $(Å^2)$	25.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The analyses of the Patterson function reveals a significant off-origin peak that is 31.93 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.0217e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.

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



<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: P6G, IOD, EDO, CL

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
	Unam	RMSZ	# Z  > 5	RMSZ	# Z  > 5
1	А	0.63	0/1128	0.76	0/1529
1	В	0.65	0/1100	0.75	0/1492
1	С	0.64	0/1118	0.76	0/1516
1	D	0.63	0/1133	0.76	0/1535
1	Ε	0.63	0/1073	0.75	0/1457
1	F	0.64	0/1109	0.76	0/1507
All	All	0.64	0/6661	0.76	0/9036

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 (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	А	1085	0	1073	9	0
1	В	1062	0	1038	6	0
1	С	1075	0	1051	3	0
1	D	1087	0	1077	1	0
1	Е	1045	0	1016	6	0
1	F	1066	0	1054	2	0
2	А	2	0	0	0	0



Mol	Chain	Non-H	- •	H(added)	Clashes	Symm-Clashes
2	Е	2	0	0	0	0
2	F	2	0	0	0	0
3	А	19	0	26	5	0
3	В	19	0	26	1	0
3	С	19	0	26	4	0
4	А	4	0	5	0	0
4	Ε	4	0	6	0	0
5	В	1	0	0	0	0
5	С	2	0	0	0	0
6	А	182	0	0	1	0
6	В	170	0	0	3	0
6	С	186	0	0	1	0
6	D	109	0	0	0	0
6	Ε	122	0	0	0	0
6	F	111	0	0	0	0
All	All	7374	0	6398	28	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 2.

The worst 5 of 28 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)
3:A:902:P6G:H142	1:E:758:ASN:HD22	1.31	0.93
3:A:902:P6G:H142	1:E:758:ASN:ND2	2.12	0.58
1:B:801:TYR:CE2	3:B:902:P6G:H61	2.44	0.51
1:A:758:ASN:HD22	3:A:902:P6G:C6	2.25	0.50
1:A:680:GLU:HG3	1:A:735:PHE:CG	2.46	0.50

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	Perce	ntiles
1	А	136/131~(104%)	132~(97%)	4(3%)	0	100	100
1	В	132/131 (101%)	127 (96%)	5(4%)	0	100	100
1	С	135/131 (103%)	131 (97%)	4 (3%)	0	100	100
1	D	136/131 (104%)	132 (97%)	4 (3%)	0	100	100
1	Ε	129/131~(98%)	125~(97%)	4(3%)	0	100	100
1	F	134/131~(102%)	130 (97%)	4 (3%)	0	100	100
All	All	802/786~(102%)	777~(97%)	25~(3%)	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	А	126/119~(106%)	124~(98%)	2(2%)	62	27
1	В	122/119 (102%)	121 (99%)	1 (1%)	81	57
1	С	125/119~(105%)	125 (100%)	0	100	100
1	D	126/119~(106%)	126 (100%)	0	100	100
1	Ε	119/119 (100%)	115~(97%)	4 (3%)	37	5
1	F	124/119~(104%)	121 (98%)	3(2%)	49	12
All	All	742/714~(104%)	732~(99%)	10 (1%)	73	34

5 of 10 residues with a non-rotameric side chain are listed below:

Mol	Chain	Res	Type
1	F	722[A]	ASN
1	F	722[B]	ASN
1	F	792	HIS
1	Е	745	ARG
1	Е	792	HIS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:



Mol	Chain	Res	Type
1	А	758	ASN
1	В	805	ASN
1	С	758	ASN
1	Е	758	ASN
1	Е	792	HIS

#### 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)

Of 14 ligands modelled in this entry, 9 are monoatomic - leaving 5 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 Turne		Chain Res		Link	Bond lengths			Bond angles		
	Mol Type Chain			Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2	
3	P6G	В	902	-	18,18,18	0.25	0	$17,\!17,\!17$	0.26	0
3	P6G	С	903	-	18,18,18	0.23	0	17,17,17	0.29	0
3	P6G	А	902	-	18,18,18	0.42	0	17,17,17	0.41	0
4	EDO	Е	903	-	3,3,3	0.13	0	2,2,2	0.10	0
4	EDO	А	903	-	3,3,3	0.60	0	2,2,2	0.28	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	P6G	В	902	-	-	4/16/16/16	-
3	P6G	С	903	-	-	10/16/16/16	-
3	P6G	А	902	-	-	8/16/16/16	-
4	EDO	Е	903	-	-	1/1/1/1	-
4	EDO	А	903	-	-	1/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 24 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	В	902	P6G	O1-C2-C3-O4
3	А	902	P6G	O10-C11-C12-O13
3	А	902	P6G	O4-C5-C6-O7
3	В	902	P6G	O7-C8-C9-O10
3	С	903	P6G	C6-C5-O4-C3

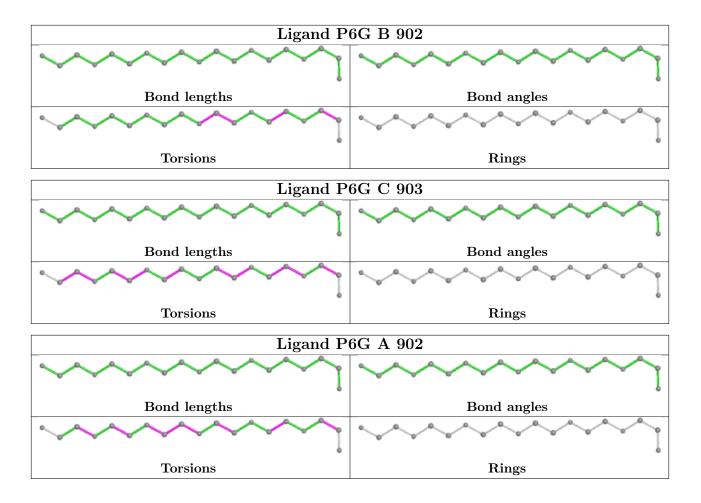
There are no ring outliers.

3 monomers are involved in 10 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	В	902	P6G	1	0
3	С	903	P6G	4	0
3	А	902	P6G	5	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.





## 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 $>$	# RSRZ > 2	$OWAB(Å^2)$	Q < 0.9
1	А	131/131~(100%)	-0.38	1 (0%) 86 86	14, 20, 33, 50	0
1	В	130/131~(99%)	-0.37	1 (0%) 86 86	15, 20, 35, 48	0
1	С	131/131 (100%)	-0.42	1 (0%) 86 86	14, 19, 30, 50	0
1	D	131/131 (100%)	-0.29	2 (1%) 73 75	15, 24, 41, 58	0
1	Ε	130/131~(99%)	0.03	8 (6%) 20 19	16, 25, 45, 53	0
1	F	130/131~(99%)	-0.28	1 (0%) 86 86	16, 25, 43, 50	0
All	All	783/786~(99%)	-0.28	14 (1%) 68 69	14, 22, 40, 58	0

The worst 5 of 14 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	Ε	729	ASP	4.1
1	F	714	ASP	4.0
1	Е	730	GLY	3.9
1	Е	731	GLU	3.5
1	Е	805	ASN	3.3

## 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.



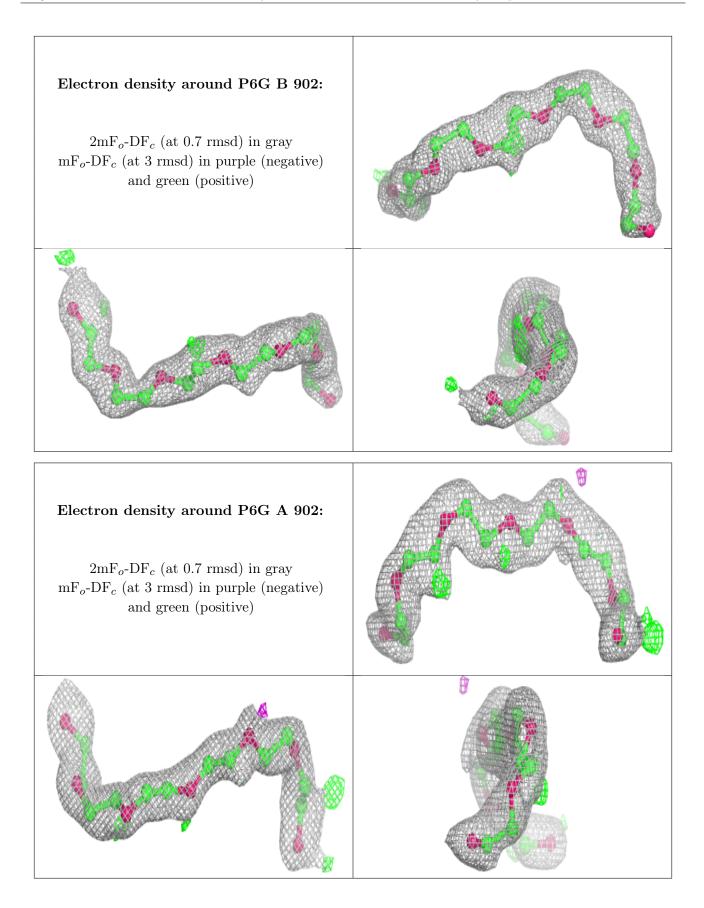
### 6.4 Ligands (i)

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.

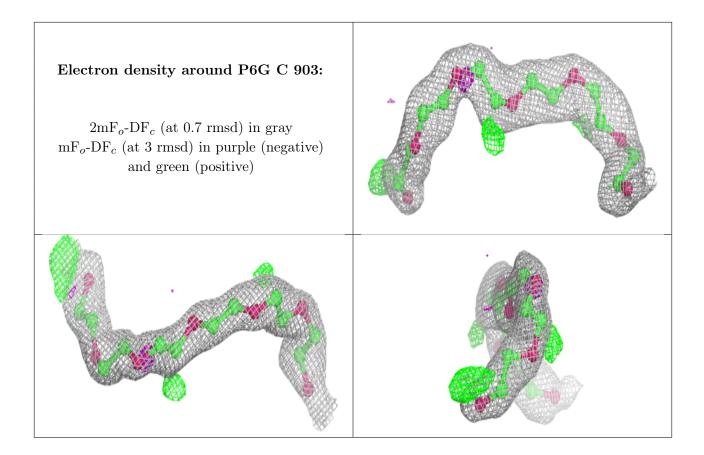
Mol	Type	Chain	Res	Atoms	RSCC	RSR	$B-factors(Å^2)$	$Q{<}0.9$
3	P6G	В	902	19/19	0.84	0.12	$35,\!39,\!46,\!49$	0
3	P6G	А	902	19/19	0.86	0.08	32,43,49,51	0
3	P6G	С	903	19/19	0.87	0.09	29,39,48,54	0
4	EDO	Е	903	4/4	0.88	0.08	40,43,44,47	0
4	EDO	А	903	4/4	0.96	0.14	20,32,36,39	0
2	CL	F	902	1/1	0.97	0.12	59, 59, 59, 59	0
2	CL	Е	901	1/1	0.97	0.03	45,45,45,45	0
2	CL	А	904	1/1	0.98	0.07	44,44,44,44	0
2	CL	А	901	1/1	0.99	0.04	31,31,31,31	0
2	CL	Е	902	1/1	0.99	0.07	$35,\!35,\!35,\!35$	0
2	CL	F	901	1/1	1.00	0.03	42,42,42,42	0
5	IOD	В	901	1/1	1.00	0.05	18,18,18,18	0
5	IOD	С	901	1/1	1.00	0.04	21,21,21,21	0
5	IOD	С	902	1/1	1.00	0.02	22,22,22,22	1

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.

