

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

#### Dec 4, 2023 - 04:19 PM JST

PDB ID	:	8HV3
Title	:	Crystal structure of EGFR_DMX in complex with covalently bound fragment
		4
Authors	:	Dokurno, P.
Deposited on	:	2022-12-26
Resolution	:	2.40  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 (1)) were used in the production of this report:

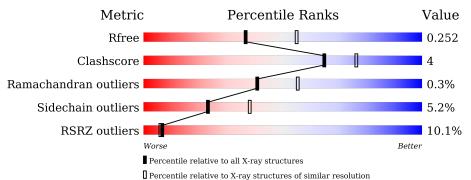
Xtriage (Phenix) EDS buster-report Percentile statistics Refmac CCP4 Ideal geometry (proteins) Ideal geometry (DNA, RNA)	:::::::::::::::::::::::::::::::::::::::	20191225.v01 (using entries in the PDB archive December 25th 2019) 5.8.0158 7.0.044 (Gargrove) Engh & Huber (2001) Parkinson et al. (1996)
Ideal geometry (DNA, RNA) Validation Pipeline (wwPDB-VP)		Parkinson et al. (1996) 2.36

# 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 2.40 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 <sub>free</sub>	130704	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

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			
			9%			
1	А	331	81%	10%	• 7	%



## 2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 2556 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.

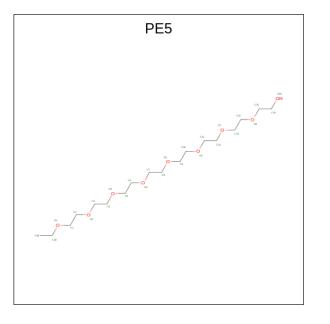
• Molecule 1 is a protein called Epidermal growth factor receptor.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
1	А	307	Total 2381	C 1526	N 401	O 436	S 18	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
А	694	GLY	-	- expression tag	
А	790	MET	THR	variant	UNP P00533
А	858	ARG	LEU	variant	UNP P00533
А	865	ALA	GLU	engineered mutation	UNP P00533
А	866	ALA	GLU	engineered mutation	UNP P00533
А	867	ALA	LYS	engineered mutation	UNP P00533
А	1023	ASN	-	expression tag	UNP P00533
А	1024	SER	-	expression tag	UNP P00533

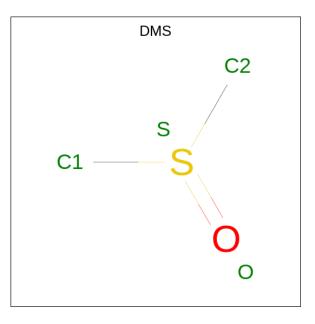
• Molecule 2 is 3,6,9,12,15,18,21,24-OCTAOXAHEXACOSAN-1-OL (three-letter code: PE5) (formula:  $C_{18}H_{38}O_9$ ).





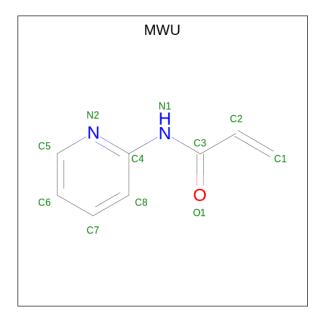
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	А	1	Total 26	C 18	0 8	0	0

• Molecule 3 is DIMETHYL SULFOXIDE (three-letter code: DMS) (formula:  $C_2H_6OS$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	А	1	Total 4	С 2	0 1	S 1	0	0

• Molecule 4 is  $\{N\}$ -pyridin-2-ylprop-2-enamide (three-letter code: MWU) (formula:  $C_8H_8N_2O$ ) (labeled as "Ligand of Interest" by depositor).





Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	А	1	Total 11	C 8	N 2	0 1	0	0

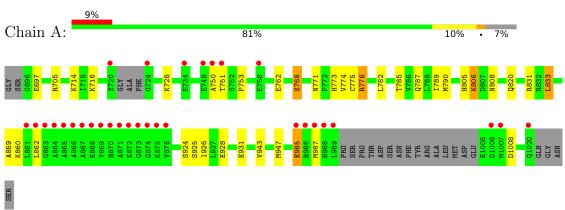
• Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	А	134	Total O 134 134	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: Epidermal growth factor receptor



## 4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 2 3	Depositor
Cell constants	145.04Å $145.04$ Å $145.04$ Å	Depositor
a, b, c, $\alpha$ , $\beta$ , $\gamma$	$90.00^{\circ}$ $90.00^{\circ}$ $90.00^{\circ}$	Depositor
Resolution (Å)	20.00 - 2.40	Depositor
	19.74 - 2.40	EDS
% Data completeness	99.3 (20.00-2.40)	Depositor
(in resolution range)	99.6(19.74-2.40)	EDS
R <sub>merge</sub>	0.16	Depositor
R <sub>sym</sub>	(Not available)	Depositor
$< I/\sigma(I) > 1$	$1.75 (at 2.41 \text{\AA})$	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
$R, R_{free}$	0.212 , $0.250$	Depositor
It, Itfree	0.216 , $0.252$	DCC
$R_{free}$ test set	977 reflections $(4.92\%)$	wwPDB-VP
Wilson B-factor ( $Å^2$ )	35.1	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent $k_{sol}(e/Å^3), B_{sol}(Å^2)$	0.37, $60.1$	EDS
L-test for twinning <sup>2</sup>	$<  L  > = 0.55, < L^2 > = 0.39$	Xtriage
Estimated twinning fraction	0.003 for -l,-k,-h	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	2556	wwPDB-VP
Average B, all atoms $(Å^2)$	48.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 3.24% of the height of the origin peak. No significant pseudotranslation is detected.

<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: PE5, MWU, DMS

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		
	Chain	RMSZ	# Z  > 5	RMSZ	# Z  > 5	
1	А	0.74	0/2429	0.89	1/3296~(0.0%)	

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$Observed(^{o})$	$Ideal(^{o})$
1	А	776	ARG	N-CA-CB	-6.10	99.62	110.60

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	2381	0	2366	19	0
2	А	26	0	35	1	0
3	А	4	0	6	0	0
4	А	11	0	0	0	0
5	А	134	0	0	3	0
All	All	2556	0	2407	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 4.



Atom-1	Atom-2	Interatomic	Clash
Atom-1	Atom-2	distance (Å)	overlap (Å)
1:A:924:SER:O	1:A:928:GLU:HG3	1.71	0.91
2:A:1101:PE5:C16	5:A:1273:HOH:O	2.35	0.73
1:A:762:GLU:OE2	5:A:1201:HOH:O	2.15	0.65
1:A:771:ASN:HB3	1:A:774:VAL:HG22	1.80	0.62
1:A:833:LEU:HD21	1:A:859:ALA:HB1	1.85	0.58
1:A:750:ALA:HB2	1:A:785:THR:HG22	1.91	0.51
1:A:924:SER:O	1:A:928:GLU:CG	2.53	0.51
1:A:985:GLU:H	1:A:985:GLU:CD	2.16	0.49
1:A:833:LEU:CD2	1:A:859:ALA:HB1	2.41	0.49
1:A:714:LYS:HE2	1:A:787:GLN:HE22	1.78	0.48
1:A:789:ILE:HD12	1:A:789:ILE:N	2.28	0.48
1:A:943:VAL:HG12	1:A:947:MET:HE2	1.95	0.48
1:A:926:ILE:HG23	1:A:931:GLU:HB3	1.95	0.47
1:A:768:SER:O	1:A:831:ARG:NH2	2.34	0.47
1:A:751:THR:O	1:A:753:PRO:HD3	2.17	0.45
1:A:859:ALA:O	1:A:860:LYS:HE3	2.17	0.45
1:A:773:HIS:CE1	1:A:820:GLN:HG2	2.53	0.44
1:A:1008:ASP:N	5:A:1211:HOH:O	2.53	0.42
1:A:716:LYS:CG	1:A:728:LYS:HE3	2.50	0.41
1:A:805:HIS:O	1:A:806:LYS:C	2.58	0.41

All (20) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

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	А	301/331~(91%)	289~(96%)	11 (4%)	1 (0%)	41	55

All (1) Ramachandran outliers are listed below:



Mol	Chain	Res	Type
1	А	862	LEU

#### 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 side chain conformation was analysed, and the total number of residues.

Mo	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	А	252/287~(88%)	239~(95%)	13~(5%)	23 38	

All (13) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	А	697	GLU
1	А	705	ARG
1	А	768	SER
1	А	775	CYS
1	А	776	ARG
1	А	782	LEU
1	А	790	MET
1	А	806	LYS
1	А	808	ASN
1	А	833	LEU
1	А	925	SER
1	А	985	GLU
1	А	987	MET

Sometimes side chains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such side chains are listed below:

Mol	Chain	Res	Type
1	А	787	GLN
1	А	812	GLN
1	А	888	HIS
1	А	976	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 monosaccharides in this entry.

#### 5.6 Ligand geometry (i)

3 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).

Mal	Mol Type Chain I		Res	Link	Bond lengths			Bond angles		
	туре	Unam	nes		Counts	RMSZ	# Z >2	Counts	RMSZ	# Z  > 2
2	PE5	А	1101	-	25,25,26	0.43	0	24,24,25	0.22	0
3	DMS	А	1102	-	3,3,3	0.38	0	3,3,3	0.31	0
4	MWU	А	1103	1	11,11,11	1.48	1 (9%)	13,13,13	3.08	3 (23%)

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	PE5	А	1101	-	-	15/23/23/24	-
4	MWU	А	1103	1	-	2/6/6/6	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Ζ	Observed(Å)	Ideal(Å)
4	A	1103	MWU	C1-C2	4.61	1.53	1.30

All (3) bond angle outliers are listed below:



Mol	Chain	Res	Type	Atoms	Ζ	$Observed(^{o})$	$Ideal(^{o})$
4	А	1103	MWU	C1-C2-C3	-8.25	112.83	122.27
4	А	1103	MWU	C2-C3-N1	6.29	117.98	113.84
4	А	1103	MWU	O1-C3-N1	-3.54	118.99	123.05

There are no chirality outliers.

All (17) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	А	1103	MWU	C2-C3-N1-C4
2	А	1101	PE5	O1-C1-C2-O2
4	А	1103	MWU	O1-C3-N1-C4
2	А	1101	PE5	O2-C3-C4-O3
2	А	1101	PE5	O4-C7-C8-O5
2	А	1101	PE5	O7-C13-C14-O8
2	А	1101	PE5	C48-C50-O1-C1
2	А	1101	PE5	C6-C5-O3-C4
2	А	1101	PE5	C8-C7-O4-C6
2	А	1101	PE5	C16-C15-O8-C14
2	А	1101	PE5	O3-C5-C6-O4
2	А	1101	PE5	C4-C3-O2-C2
2	А	1101	PE5	C9-C10-O6-C11
2	А	1101	PE5	C14-C13-O7-C12
2	А	1101	PE5	C7-C8-O5-C9
2	А	1101	PE5	C1-C2-O2-C3
2	А	1101	PE5	C11-C12-O7-C13

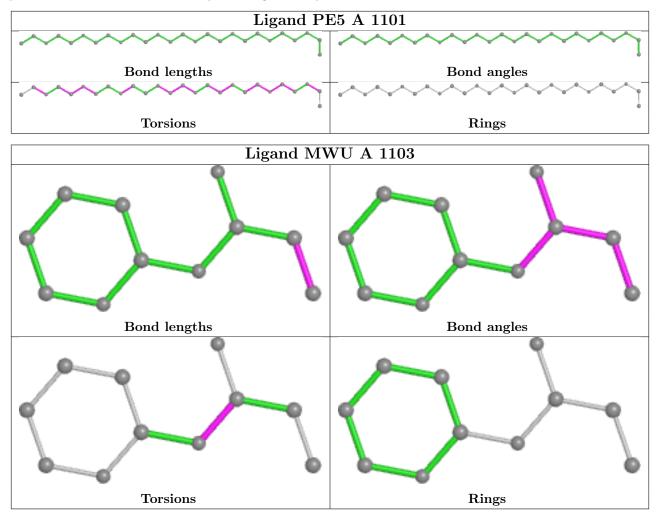
There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	$\mathbf{Res}$	Type	Clashes	Symm-Clashes
2	А	1101	PE5	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.

### 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	$\mathbf{OWAB}(\mathrm{\AA}^2)$	Q < 0.9
1	А	307/331~(92%)	0.25	31 (10%) 7 6	22, 40, 112, 184	0

All (31) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	А	863	GLY	12.3
1	А	864	ALA	9.8
1	А	866	ALA	9.7
1	А	865	ALA	9.0
1	А	869	TYR	8.7
1	А	862	LEU	7.2
1	А	751	THR	7.0
1	А	872	GLU	6.7
1	А	867	ALA	6.0
1	А	875	LYS	5.8
1	А	874	GLY	5.7
1	А	873	GLY	5.5
1	А	871	ALA	5.4
1	А	870	HIS	5.3
1	А	988	HIS	4.1
1	А	987	MET	4.0
1	А	724	GLY	3.6
1	А	986	ARG	3.4
1	А	989	LEU	3.2
1	А	750	ALA	2.8
1	А	749	GLU	2.7
1	А	861	LEU	2.7
1	А	1007	MET	2.6
1	А	985	GLU	2.5
1	А	734	GLU	2.5
1	А	868	GLU	2.4
1	А	1020	GLN	2.4

Continued on next page...



Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	А	758	GLU	2.3
1	А	720	SER	2.2
1	А	1006	ASP	2.2
1	А	876	VAL	2.1

#### 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 monosaccharides in this entry.

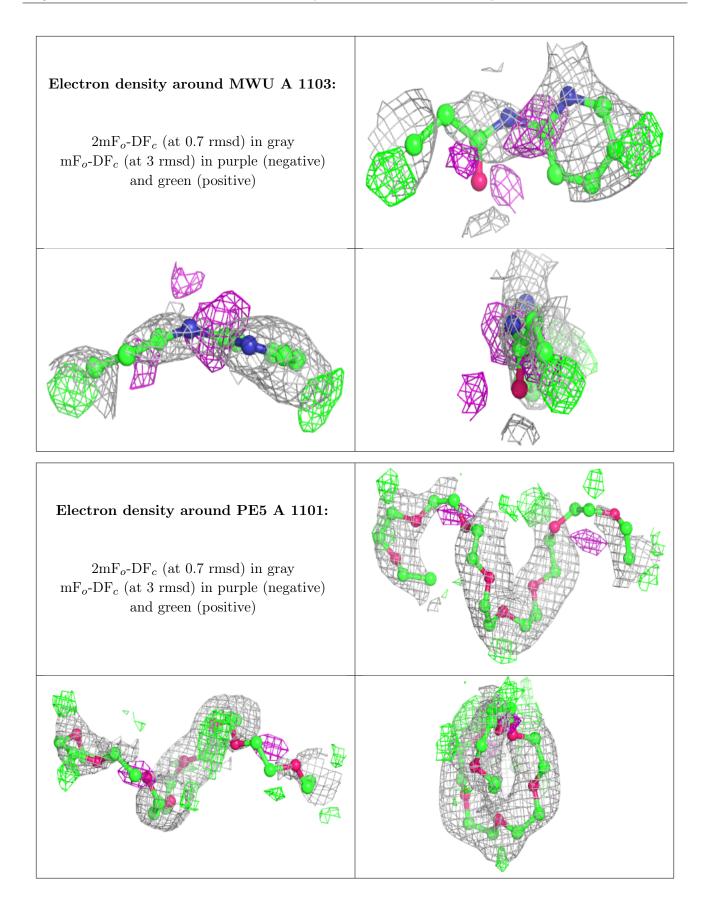
#### 6.4 Ligands (i)

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	$\operatorname{B-factors}(\operatorname{\AA}^2)$	Q<0.9
4	MWU	А	1103	11/11	0.78	0.32	47,62,64,70	0
2	PE5	А	1101	26/27	0.83	0.24	32,45,94,100	26
3	DMS	А	1102	4/4	0.90	0.17	67,78,84,88	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.

