



Full wwPDB X-ray Structure Validation Report ⓘ

May 28, 2024 – 07:43 pm BST

PDB ID : 8PO2
Title : Discovery and Optimisation of Potent, Efficacious and Selective Inhibitors Targeting EGFR Exon20 Insertion Mutations. Compound 33 bound to EGFRin-sNPG [V948R]
Authors : Hargreaves, D.
Deposited on : 2023-07-03
Resolution : 2.28 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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.8.4, CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.36.2
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) : 2.36.2

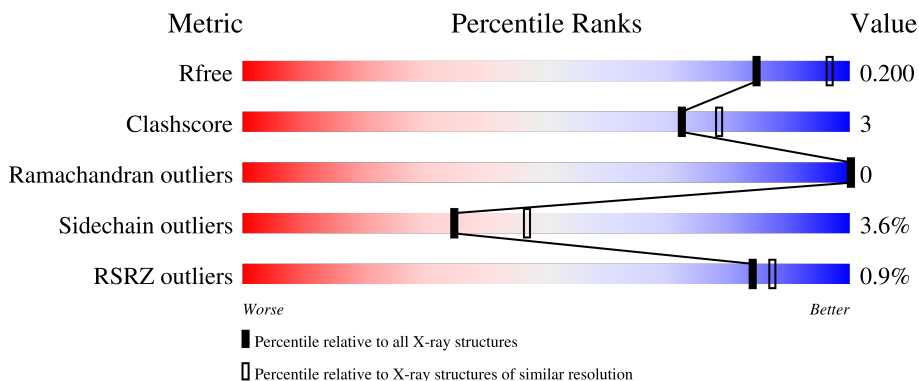
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION




The reported resolution of this entry is 2.28 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	6980 (2.30-2.26)
Clashscore	141614	7711 (2.30-2.26)
Ramachandran outliers	138981	7597 (2.30-2.26)
Sidechain outliers	138945	7598 (2.30-2.26)
RSRZ outliers	127900	6849 (2.30-2.26)

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 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 $\leq 5\%$. 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	332	 73% 9% 18%
1	B	332	 72% 10% 17%
1	D	332	 5% 5% 5% 5% 95%

2 Entry composition i

There are 3 unique types of molecules in this entry. The entry contains 4810 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.

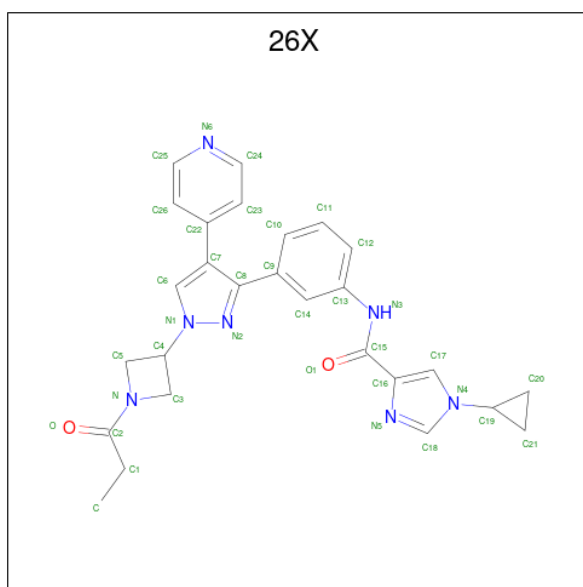
- Molecule 1 is a protein called Epidermal growth factor receptor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	271	Total 2172	C 1401	N 371	O 385	S 15	0	0	0
1	B	274	Total 2193	C 1416	N 374	O 388	S 15	0	0	0
1	D	18	Total 145	C 88	N 18	O 37	S 2	0	0	0

There are 15 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	694	GLY	-	expression tag	UNP P00533
A	773	GLY	-	insertion	UNP P00533
A	774	ASN	-	insertion	UNP P00533
A	775	PRO	-	insertion	UNP P00533
A	951	ARG	VAL	engineered mutation	UNP P00533
B	694	GLY	-	expression tag	UNP P00533
B	773	GLY	-	insertion	UNP P00533
B	774	ASN	-	insertion	UNP P00533
B	775	PRO	-	insertion	UNP P00533
B	951	ARG	VAL	engineered mutation	UNP P00533
D	-311	GLY	-	expression tag	UNP P00533
D	-232	GLY	-	insertion	UNP P00533
D	-231	ASN	-	insertion	UNP P00533
D	-230	PRO	-	insertion	UNP P00533
D	-54	ARG	VAL	engineered mutation	UNP P00533

- Molecule 2 is 1-cyclopropyl- {N}-[3-[1-(1-propanoylazetid-3-yl)-4-pyridin-4-yl-pyrazol-3-yl]phenyl]imidazole-4-carboxamide (three-letter code: 26X) (formula: C₂₇H₂₇N₇O₂) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
2	A	1	36	27	7	2	0	0
2	B	1	36	27	7	2	0	0

- Molecule 3 is water.

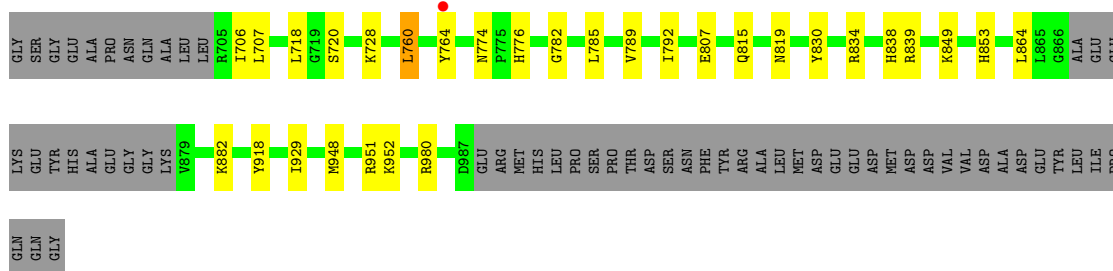
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	109	109	109	0	0
3	B	108	108	108	0	0
3	D	11	11	11	0	0

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.

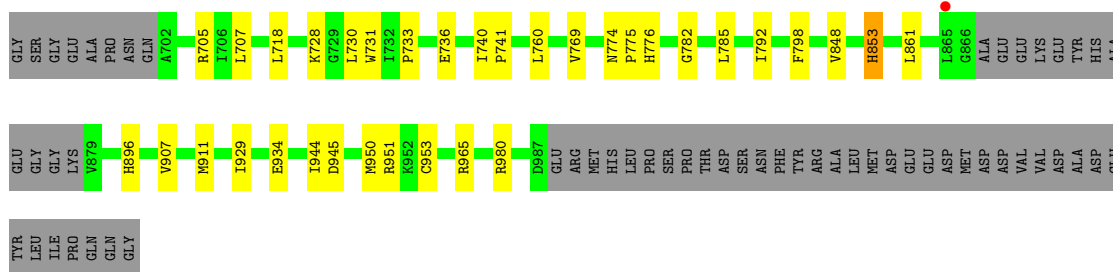
- Molecule 1: Epidermal growth factor receptor

Chain A:  73% 9% 18%



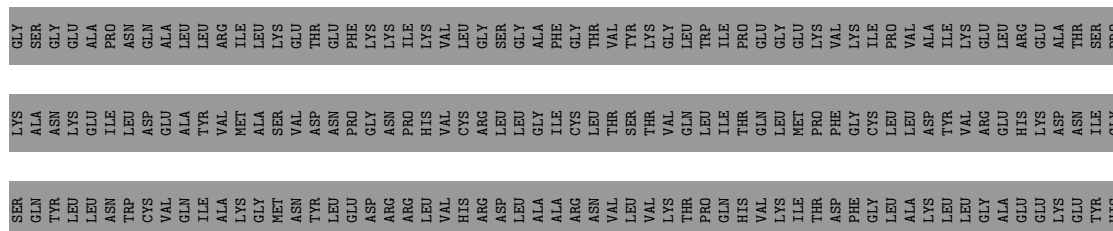
- Molecule 1: Epidermal growth factor receptor

Chain B:  72% 10% 17%



- Molecule 1: Epidermal growth factor receptor

Chain D:  5% 95%



ALA
GLU
GLY
LEU
GLY
GLY
PRO
GLN
LYS
VAL
PRO
ILE
LYS
THR
MET
ALA
LEU
GLU
SER

GLU
ARG
LEU
PRO
GLN
PRO
ILE
CYS
THR
ILE
ASP
VAL
TYR
MET
ILE
MET
ARG
LYS
CYS
TRP
MET
ILE
ASP
ALA
ASP
SER
ARG
TRP
PRO
LYS
PHE
ARG
GLU
LEU
ILE
ILE
GLU
GLY
PHE
SER
LYS
MET
ALA
ARG
ASP
PRO
GLN
ARG
TYR
ILE
LEU
VAL
ILE
GLN
GLY
ASP
SER
GLY
ARG
MET
HIS
LEU
LYS
PRO
GLY

SER
PRO
THR
ASP
SER
ASN
PHE
TYR
ARG
ALA
LEU
HO
D7
V8
V9
D12
E13
Y14
F17
GLN
GLN
GLY

4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	137.12Å 85.71Å 97.84Å 90.00° 121.76° 90.00°	Depositor
Resolution (Å)	69.06 – 2.28 69.06 – 2.27	Depositor EDS
% Data completeness (in resolution range)	70.2 (69.06-2.28) 70.2 (69.06-2.27)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.74 (at 2.27Å)	Xtrriage
Refinement program	BUSTER 2.11.8 (24-FEB-2021)	Depositor
R, R_{free}	0.194 , 0.246 0.191 , 0.200	Depositor DCC
R_{free} test set	1539 reflections (4.92%)	wwPDB-VP
Wilson B-factor (Å ²)	41.2	Xtrriage
Anisotropy	0.010	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 38.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	4810	wwPDB-VP
Average B, all atoms (Å ²)	44.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.04% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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.

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: 26X

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	0.43	0/2220	0.59	0/3003
1	B	0.41	0/2241	0.59	0/3032
1	D	0.40	0/146	0.64	0/198
All	All	0.42	0/4607	0.59	0/6233

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	A	2172	0	2235	15	0
1	B	2193	0	2262	16	0
1	D	145	0	123	2	0
2	A	36	0	0	0	0
2	B	36	0	0	1	0
3	A	109	0	0	0	0
3	B	108	0	0	1	0
3	D	11	0	0	1	0
All	All	4810	0	4620	31	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 3.

All (31) 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:A:706:ILE:HD11	1:A:760:LEU:HD11	1.76	0.66
1:B:896:HIS:HD2	3:B:1253:HOH:O	1.79	0.65
1:B:798:PHE:HB2	1:B:848:VAL:HB	1.87	0.56
1:B:776:HIS:HE1	3:D:106:HOH:O	1.87	0.56
1:A:853:HIS:CE1	1:D:9:VAL:HG12	2.41	0.56
1:B:775:PRO:O	1:B:853:HIS:HE1	1.89	0.55
1:B:731:TRP:CZ2	1:B:733:PRO:HB3	2.43	0.54
1:B:929:ILE:HG23	1:B:934:GLU:HB2	1.91	0.52
1:A:706:ILE:CD1	1:A:760:LEU:HD11	2.41	0.51
1:B:774:ASN:ND2	1:B:776:HIS:H	2.09	0.49
1:A:882:LYS:HG2	1:A:918:TYR:HD2	1.77	0.49
1:A:707:LEU:HD13	1:A:782:GLY:HA3	1.96	0.47
1:A:774:ASN:HD22	1:A:776:HIS:H	1.62	0.46
1:B:907:VAL:O	1:B:911:MET:HG2	2.17	0.45
1:A:838:HIS:O	1:A:839:ARG:HB2	2.18	0.44
1:B:707:LEU:HD13	1:B:782:GLY:HA3	1.99	0.43
1:A:948:MET:O	1:A:952:LYS:HG3	2.18	0.43
1:B:950:MET:O	1:B:953:CYS:HB2	2.17	0.43
1:B:861:LEU:HD13	2:B:1101:26X:C18	2.48	0.43
1:A:718:LEU:HD21	1:A:728:LYS:HB2	2.02	0.42
1:B:760:LEU:HG	1:B:785:LEU:HD12	2.00	0.42
1:A:815:GLN:HE21	1:A:819:ASN:ND2	2.19	0.41
1:A:834:ARG:N	1:A:834:ARG:HD2	2.35	0.41
1:A:849:LYS:HZ3	1:D:7:ASP:HB3	1.85	0.41
1:A:830:TYR:CE1	1:A:834:ARG:HD3	2.56	0.41
1:B:740:ILE:HA	1:B:741:PRO:HD3	2.00	0.41
1:B:944:ILE:HD12	1:B:944:ILE:HA	1.89	0.41
1:B:707:LEU:HD22	1:B:792:ILE:HD13	2.03	0.40
1:B:718:LEU:HD21	1:B:728:LYS:HB2	2.03	0.40
1:A:707:LEU:HD22	1:A:792:ILE:HD13	2.03	0.40
1:A:785:LEU:HD22	1:A:789:VAL:HG22	2.04	0.40

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	267/332 (80%)	260 (97%)	7 (3%)	0	100	100
1	B	270/332 (81%)	261 (97%)	9 (3%)	0	100	100
1	D	16/332 (5%)	14 (88%)	2 (12%)	0	100	100
All	All	553/996 (56%)	535 (97%)	18 (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	A	240/290 (83%)	232 (97%)	8 (3%)	38	51
1	B	242/290 (83%)	233 (96%)	9 (4%)	34	45
1	D	17/290 (6%)	16 (94%)	1 (6%)	19	24
All	All	499/870 (57%)	481 (96%)	18 (4%)	35	47

All (18) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	720	SER
1	A	760	LEU
1	A	764	TYR
1	A	807	GLU
1	A	864	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	929	ILE
1	A	951	ARG
1	A	980	ARG
1	B	705	ARG
1	B	730	LEU
1	B	736	GLU
1	B	769	VAL
1	B	853	HIS
1	B	945	ASP
1	B	951	ARG
1	B	965	ARG
1	B	980	ARG
1	D	0	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (16) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	756	ASN
1	A	774	ASN
1	A	776	HIS
1	A	794	GLN
1	A	819	ASN
1	A	845	ASN
1	A	938	GLN
1	A	979	GLN
1	B	756	ASN
1	B	774	ASN
1	B	776	HIS
1	B	819	ASN
1	B	845	ASN
1	B	853	HIS
1	B	938	GLN
1	B	979	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](#)

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	26X	B	1101	1	32,41,41	0.93	2 (6%)	40,59,59	0.84	2 (5%)
2	26X	A	1101	1	32,41,41	0.72	1 (3%)	40,59,59	0.82	1 (2%)

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	26X	B	1101	1	-	5/16/40/40	0/6/6/6
2	26X	A	1101	1	-	2/16/40/40	0/6/6/6

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	1101	26X	C4-N1	-3.57	1.43	1.49
2	B	1101	26X	C8-N2	-2.61	1.33	1.35
2	A	1101	26X	C8-N2	-2.56	1.33	1.35

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	1101	26X	C5-C4-N1	-2.39	111.42	117.11
2	B	1101	26X	C6-N1-C4	2.13	127.37	125.48
2	A	1101	26X	C6-N1-C4	2.02	127.27	125.48

There are no chirality outliers.

All (7) torsion outliers are listed below:

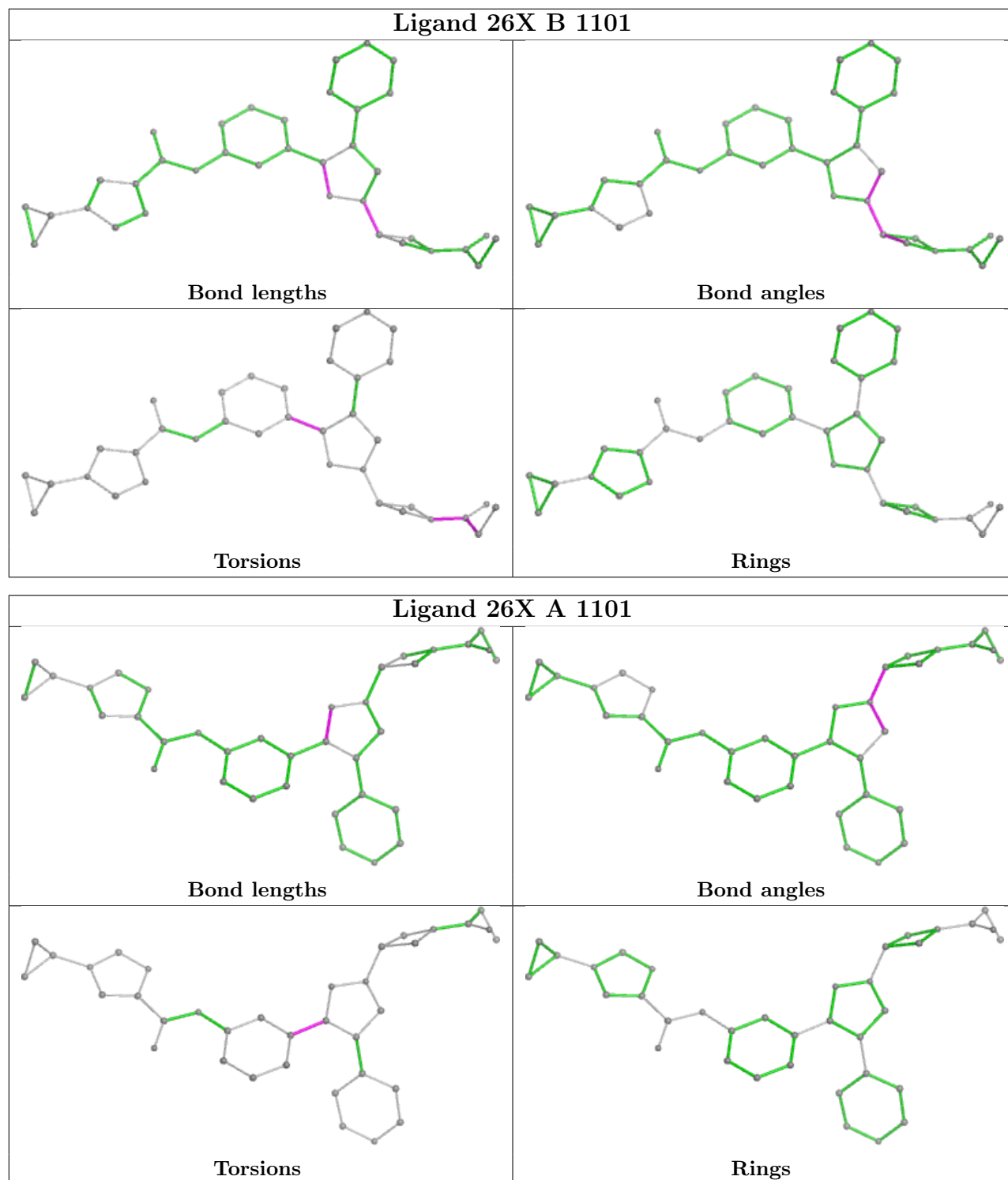
Mol	Chain	Res	Type	Atoms
2	B	1101	26X	C-C1-C2-N
2	B	1101	26X	O-C2-N-C5
2	B	1101	26X	N2-C8-C9-C10
2	B	1101	26X	N2-C8-C9-C14
2	A	1101	26X	N2-C8-C9-C10
2	A	1101	26X	N2-C8-C9-C14
2	B	1101	26X	C-C1-C2-O

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	1101	26X	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 than 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

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(Å ²)	Q<0.9
1	A	271/332 (81%)	0.02	1 (0%) 92 94	25, 43, 63, 73	0
1	B	274/332 (82%)	-0.05	1 (0%) 92 94	25, 42, 65, 82	0
1	D	18/332 (5%)	0.63	3 (16%) 1 2	37, 59, 79, 79	0
All	All	563/996 (56%)	0.00	5 (0%) 84 87	25, 43, 66, 82	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	764	TYR	4.0
1	D	0	MET	3.7
1	D	14	TYR	2.0
1	B	865	LEU	2.0
1	D	12	ASP	2.0

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](#)

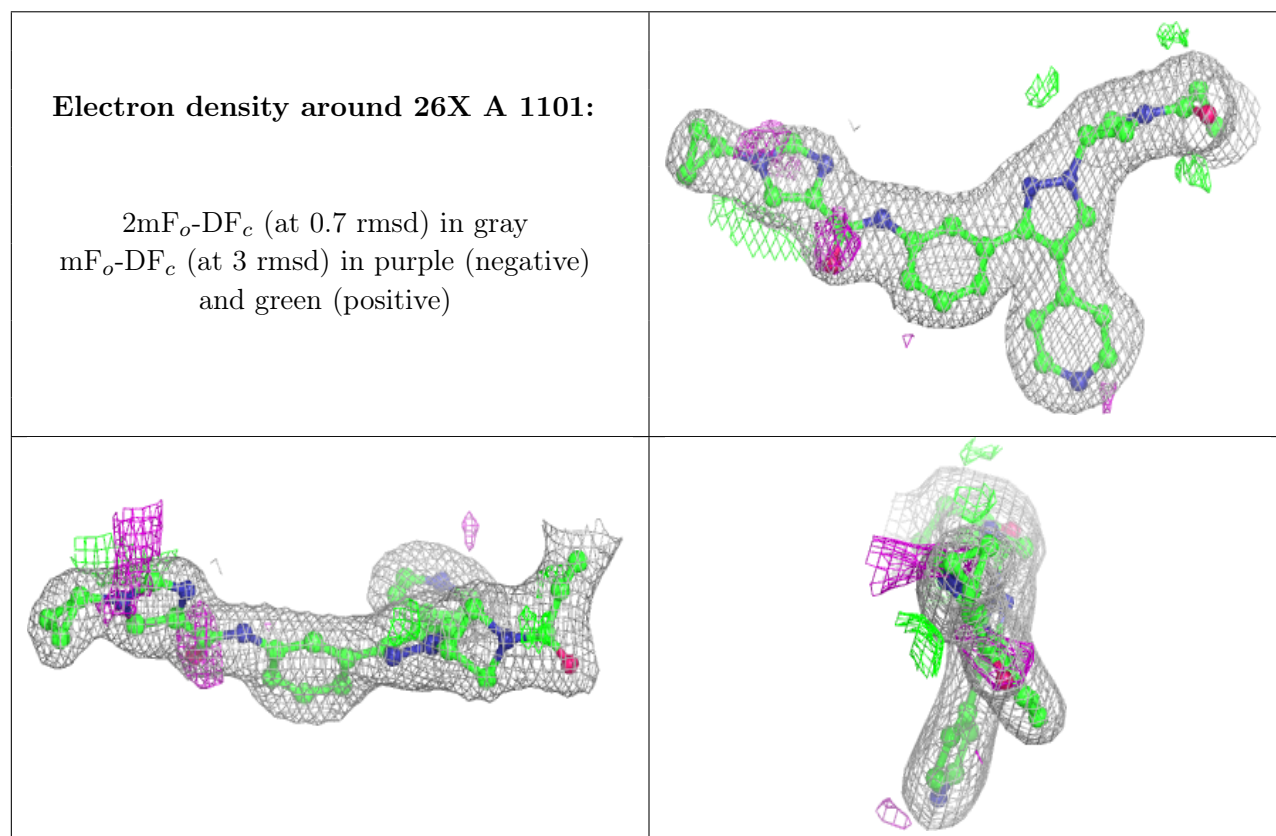
There are no monosaccharides in this entry.

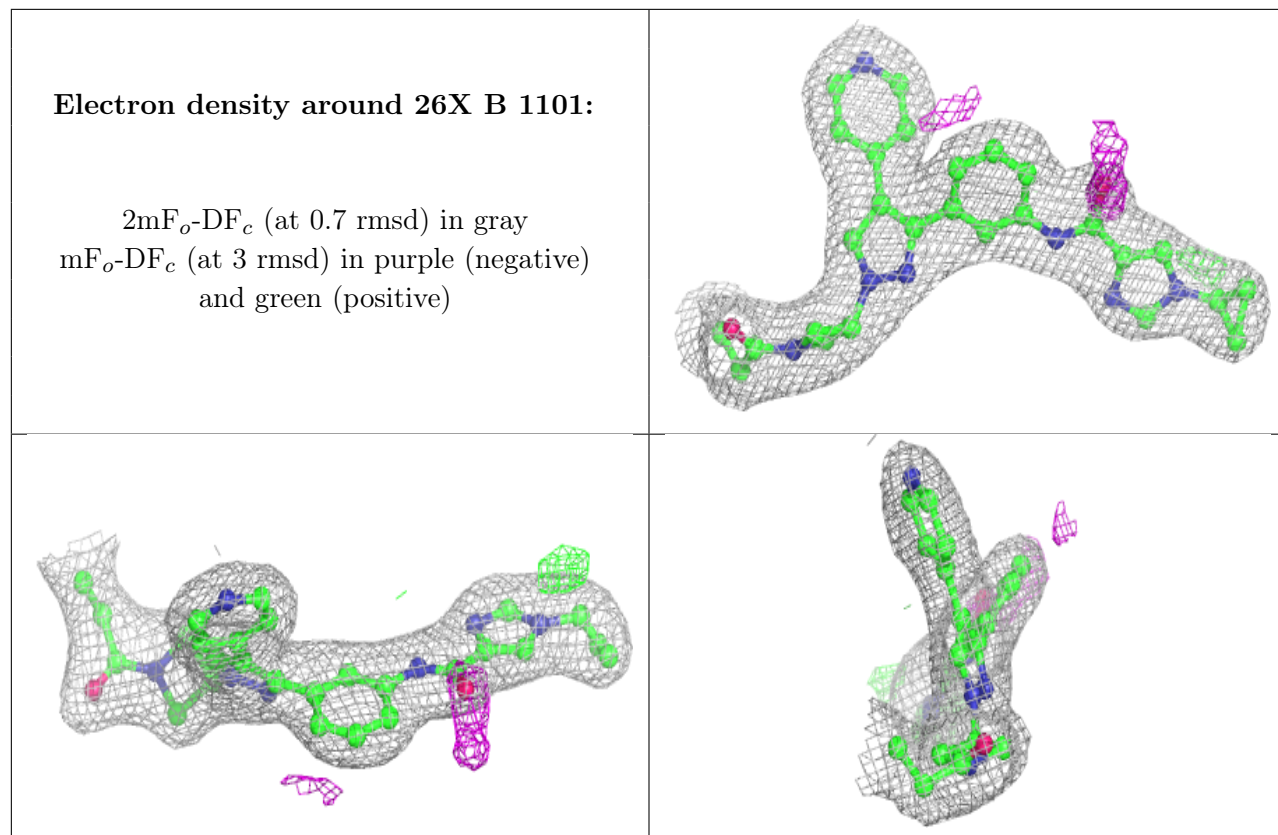
6.4 Ligands [i](#)

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, 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(\AA^2)	Q<0.9
2	26X	A	1101	36/36	0.95	0.15	33,36,47,47	0
2	26X	B	1101	36/36	0.97	0.14	36,40,44,44	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.