



# wwPDB X-ray Structure Validation Summary Report ⓘ

Apr 27, 2024 – 11:35 pm BST

PDB ID : 2VG6  
Title : Crystal structures of HIV-1 reverse transcriptase complexes with thiocarba-  
mate non-nucleoside inhibitors  
Authors : Spallarossa, A.; Cesarini, S.; Ranise, A.; Ponassi, M.; Unge, T.; Bolognesi, M.  
Deposited on : 2007-11-08  
Resolution : 3.01 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](mailto: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)  
Xtrriage (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

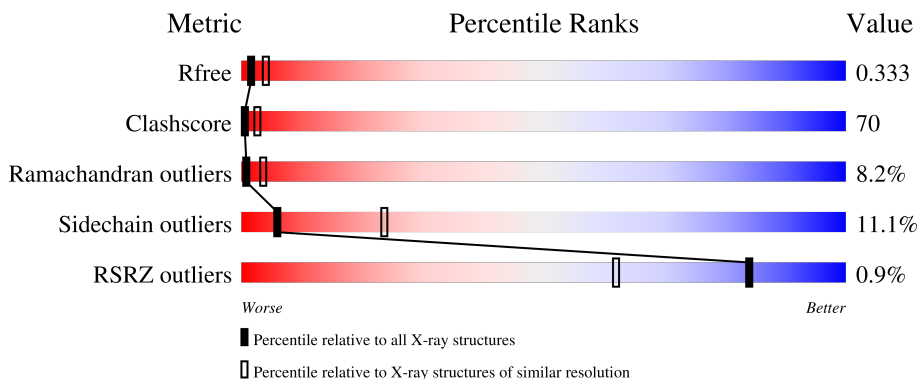
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*


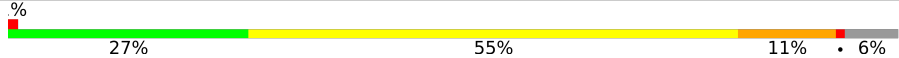
The reported resolution of this entry is 3.01 Å.

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	2399 (3.04-3.00)
Clashscore	141614	2734 (3.04-3.00)
Ramachandran outliers	138981	2640 (3.04-3.00)
Sidechain outliers	138945	2643 (3.04-3.00)
RSRZ outliers	127900	2287 (3.04-3.00)

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  $\geq 3$ , 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	557	 31% 53% 15%
2	B	428	 27% 55% 11% 6%

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 7856 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.

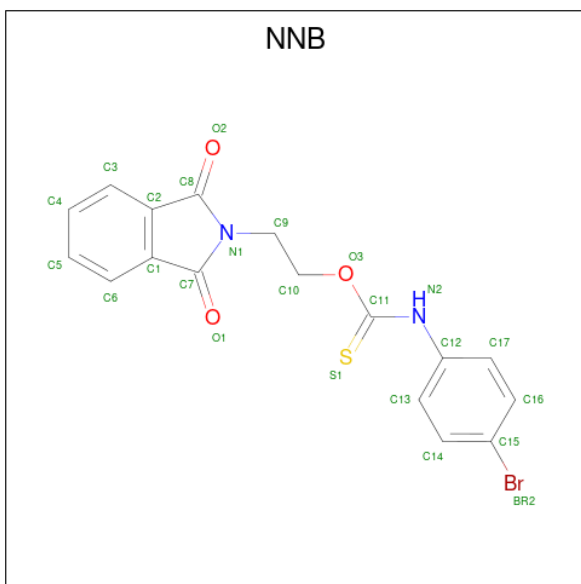
- Molecule 1 is a protein called REVERSE TRANSCRIPTASE/RIBONUCLEASE H.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	549	4468	2893	745	822	8	0	0	1

- Molecule 2 is a protein called P51 RT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	401	3318	2163	543	605	7	0	0	0

- Molecule 3 is O-[2-(1,3-dioxo-1,3-dihydro-2H-isoindol-2-yl)ethyl] (4-bromophenyl)thiocarbamate (three-letter code: NNB) (formula: C<sub>17</sub>H<sub>13</sub>BrN<sub>2</sub>O<sub>3</sub>S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	
			Total	Br	C	N	O			S
3	A	1	24	1	17	2	3	1	0	0

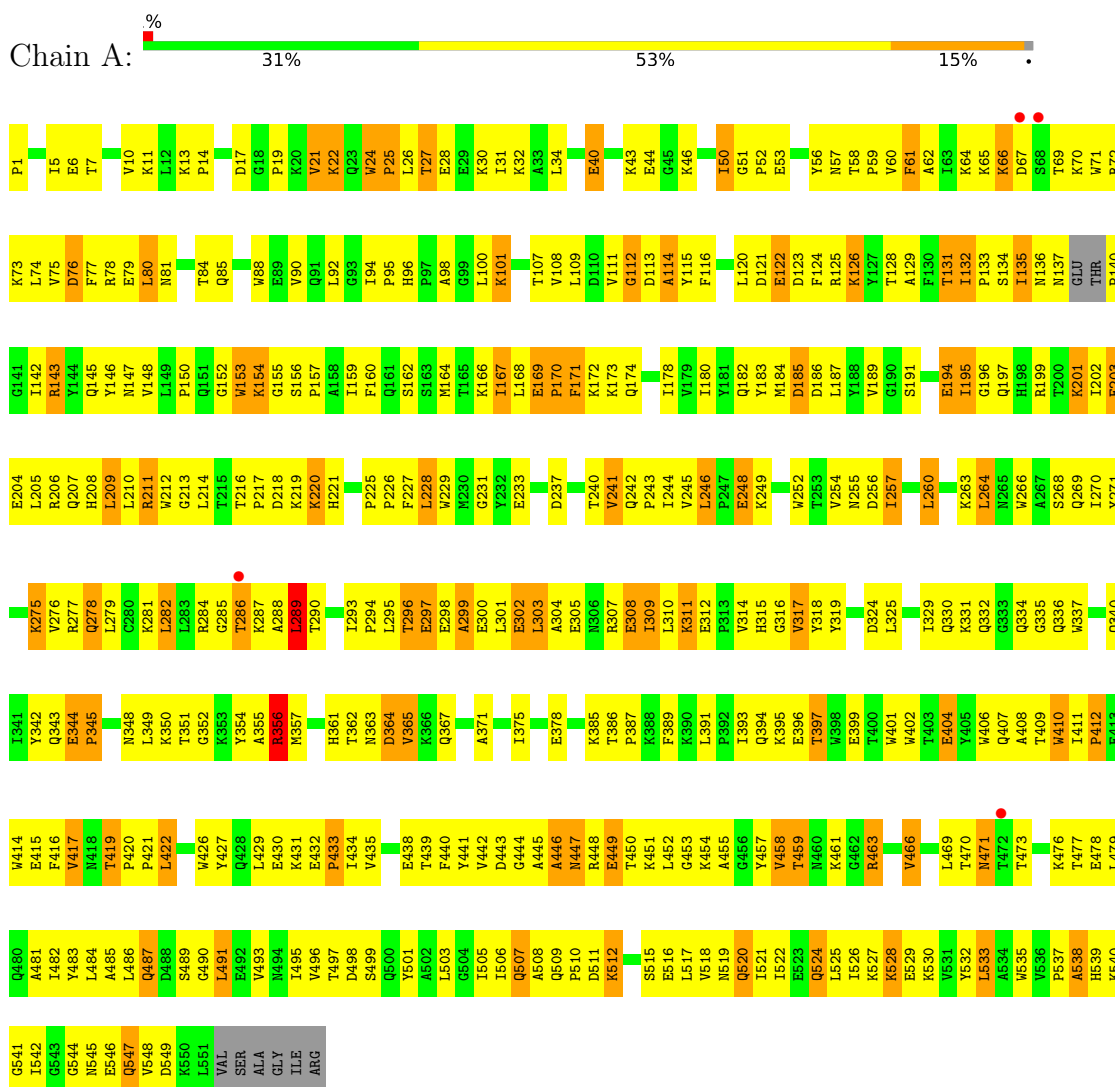
- Molecule 4 is water.

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
4	A	22	Total 22	O 22	0	0
4	B	24	Total 24	O 24	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: REVERSE TRANSCRIPTASE/RIBONUCLEASE H



#### • Molecule 2: P51 RT



E1399	W1400	W1401	W1402	T1403	E1404	Y1405	W1406	Q1407	W1410	W1414	E1415	F1416	V1417	N1418	T1419	P1420	P1421	L1422	V1423	K1424	L1425	W1426	Y1427	Q1428																																
G1333	Q1334	G1335	Q1336	W1337	T1338	Y1339	Q1340	I1341	Y1342	Q1343	E1344	F1345	K1347	M1348	L1349	K1350	T1351	G1352	A1355	ARG	MET	ARG	GLY	ALA	H1361	T1362	N1363	D1364	V1365	K1366	Q1367	E1370	K1374	I1375	T1376	T1377	E1378	S1379	I1380	V1381	I1382	W1383	G1384	K1385	K1388	F1389	K1390	L1391	P1392	I1393	Q1394	K1395	W1398			
I1270	Y1271	P1272	G1273	I1274	K1275	V1276	R1277	Q1278	L1279	C1280	K1281	L1282	LEU	ARG	GLY	T1286	K1287	A1288	L1289	T1290	E1291	V1292	I1293	P1294	L1295	T1296	E1297	E1298	A1299	E1300	L1301	E1302	L1303	A1304	E1305	N1306	R1307	E1308	I1309	L1310	K1311	S1312	P1313	V1314	V1317	D1320	P1321	S1322	K1323	D1324	L1325	I1326	I1329	Q1330	K1331	Q1332
L1209	L1210	R1211	W1212	G1213	L1214	THR	THR	PRO	ASP	LYS	LYS	HIS	GLN	LYS	GLU	PRO	PRO	PHE	LEU	TRP	M1230	G1231	Y1232	E1233	L1234	H1235	P1236	W1239	T1240	V1241	Q1242	P1243	I1244	V1245	L1246	P1247	E1248	D1250	S1251	W1252	T1253	V1254	M1255	D1256	I1257	Q1258	K1259	L1260	V1261	G1262	L1263	W1265	W1266	S1268	Q1269	
Q1145	Y1146	M1147	V1148	L1149	P1150	Q1151	G1152	W1153	K1154	G1155	S1156	P1157	A1158	I1159	F1160	Q1161	S1163	M1164	I1167	L1168	E1169	P1170	F1171	K1172	K1173	L1174	M1175	P1176	D1177	I1178	V1179	Q1182	Y1183	M1184	D1186	L1187	Y1188	L1189	G1190	S1191	D1192	L1193	I1194	I1195	G1196	Q1197	H1198	R1199	T1200	K1201	L1202	E1203	R1206			
D1067	S1068	T1069	K1073	R1074	V1075	D1076	F1077	R1078	E1079	L1080	N1081	K1082	R1083	M1084	Q1085	S1086	F1087	W1088	E1089	V1090	Q1091	L1092	G1093	I1094	V1111	G1112	D1113	A1114	Y1115	F1116	S1117	V1118	P1119	L1120	D1121	E1122	D1123	F1124	R1125	K1126	Y1130	M1135	N1136	I1142	R1143	Y1144										
PRD	I1E	SER	PRD	E1006	K1073	T1007	V1008	P1009	W1010	K1011	L1012	K1013	P1014	G1015	M1016	D1017	G1018	V1021	W1024	P1025	L1026	T1027	E1028	E1029	K1030	I1031	K1032	A1033	L1034	V1035	I1037	E1036	C1038	T1039	E1040	M1041	E1042	K1043	E1044	G1045	K1046	I1047	S1048	K1049	E1050	G1051	P1052	E1053	Y1056	M1057	Y1060	K1064	K1065	K1066		

## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	119.59Å 157.18Å 154.92Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 3.01 19.93 – 3.01	Depositor EDS
% Data completeness (in resolution range)	83.8 (20.00-3.01) 83.8 (19.93-3.01)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.31 (at 3.04Å)	Xtrriage
Refinement program	REFMAC 5.2.0005	Depositor
R, $R_{free}$	0.244 , 0.335 0.237 , 0.333	Depositor DCC
$R_{free}$ test set	1236 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	60.1	Xtrriage
Anisotropy	0.068	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 58.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	7856	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	41.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.44% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<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.

## 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: NNB

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	0.28	1/4585 (0.0%)	0.43	0/6226
2	B	0.24	0/3411	0.41	0/4632
All	All	0.26	1/7996 (0.0%)	0.42	0/10858

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

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	478	GLU	CD-OE2	7.34	1.33	1.25

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	419	THR	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	A	4468	0	4527	624	1
2	B	3318	0	3341	499	0
3	A	24	0	13	2	0
4	A	22	0	0	9	1
4	B	24	0	0	10	0
All	All	7856	0	7881	1099	1

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 70.

The worst 5 of 1099 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:B:1266:TRP:CH2	2:B:1346:PHE:HZ	1.25	1.49
2:B:1266:TRP:CH2	2:B:1346:PHE:CZ	1.99	1.48
1:A:450:THR:CG2	1:A:452:LEU:HD22	1.51	1.40
1:A:344:GLU:HB3	1:A:345:PRO:CD	1.49	1.36
2:B:1282:LEU:O	2:B:1287:LYS:NZ	1.67	1.26

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:289:LEU:CD2	4:A:2013:HOH:O[3_555]	2.06	0.14

## 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	545/557 (98%)	390 (72%)	110 (20%)	45 (8%)	<b>1</b> <b>3</b>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	B	393/428 (92%)	293 (75%)	68 (17%)	32 (8%)	1	4
All	All	938/985 (95%)	683 (73%)	178 (19%)	77 (8%)	1	3

5 of 77 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	14	PRO
1	A	28	GLU
1	A	114	ALA
1	A	135	ILE
1	A	153	TRP

### 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	490/497 (99%)	434 (89%)	56 (11%)	5	22
2	B	366/390 (94%)	327 (89%)	39 (11%)	6	25
All	All	856/887 (96%)	761 (89%)	95 (11%)	6	23

5 of 95 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	1030	LYS
2	B	1195	ILE
2	B	1048	SER
2	B	1161	GLN
2	B	1232	TYR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 38 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	1278	GLN
2	B	1407	GLN

Continued on next page...

*Continued from previous page...*

Mol	Chain	Res	Type
2	B	1330	GLN
2	B	1343	GLN
2	B	1428	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](#)

1 ligand is 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
3	NNB	A	1551	-	26,26,26	2.49	6 (23%)	36,36,36	3.17	12 (33%)

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. '2' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	NNB	A	1551	-	-	0/10/26/26	0/3/3/3

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	1551	NNB	C1-C7	-6.11	1.38	1.48
3	A	1551	NNB	C2-C8	-5.75	1.39	1.48
3	A	1551	NNB	C7-N1	-5.55	1.33	1.39
3	A	1551	NNB	C8-N1	-5.42	1.33	1.39
3	A	1551	NNB	C12-N2	-4.12	1.33	1.41

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	1551	NNB	C10-O3-C11	-12.38	109.05	119.11
3	A	1551	NNB	O3-C11-S1	-7.81	119.76	125.10
3	A	1551	NNB	C12-N2-C11	-5.20	120.71	130.00
3	A	1551	NNB	C1-C7-N1	5.06	109.48	105.88
3	A	1551	NNB	C2-C8-N1	4.44	109.03	105.88

There are no chirality outliers.

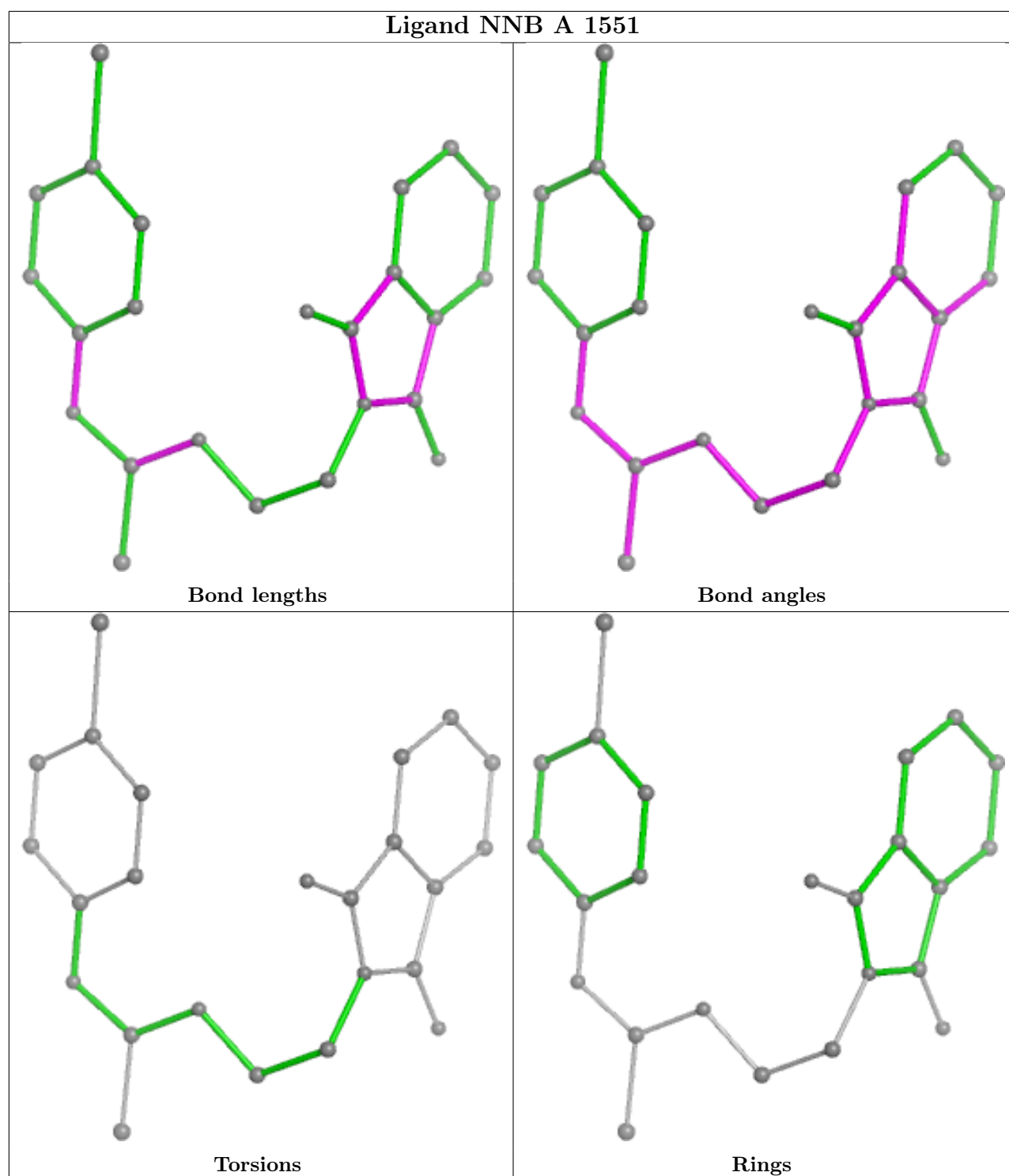
There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	1551	NNB	2	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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	549/557 (98%)	-0.42	4 (0%) 87 68	16, 41, 62, 79	0
2	B	401/428 (93%)	-0.45	5 (1%) 79 53	18, 36, 74, 85	0
All	All	950/985 (96%)	-0.43	9 (0%) 84 62	16, 40, 67, 85	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	67	ASP	3.5
2	B	1361	HIS	3.0
2	B	1069	THR	2.5
2	B	1068	SER	2.4
1	A	472	THR	2.4

### 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](#)

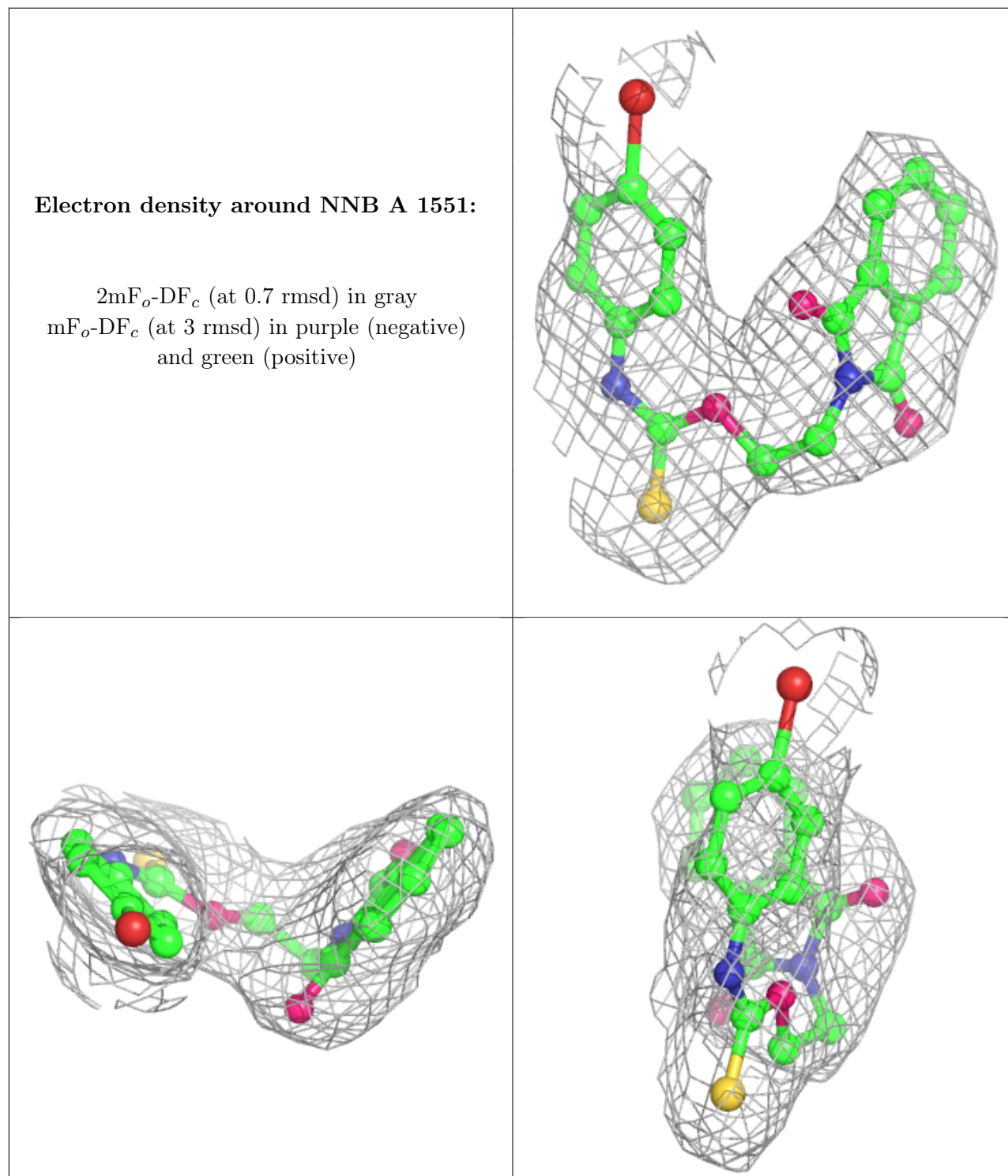
There are no monosaccharides in this entry.

### 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<sup>th</sup> 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( $\text{\AA}^2$ )	Q<0.9
3	NNB	A	1551	24/24	0.98	0.14	23,27,40,44	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.