



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 05:57 AM GMT

PDB ID : 2VG5
Title : CRYSTAL STRUCTURES OF HIV-1 REVERSE TRANSCRIPTASE COMPLEXES WITH THIOCARBAMATE NON-NUCLEOSIDE INHIBITORS
Authors : Spallarossa, A.; Cesarini, S.; Ranise, A.; Ponassi, M.; Unge, T.; Bolognesi, M.
Deposited on : 2007-11-08
Resolution : 2.80 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
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>
with specific help available everywhere you see the ⓘ symbol.

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

MolProbity : 4.02b-467
Mogul : 1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : trunk26865

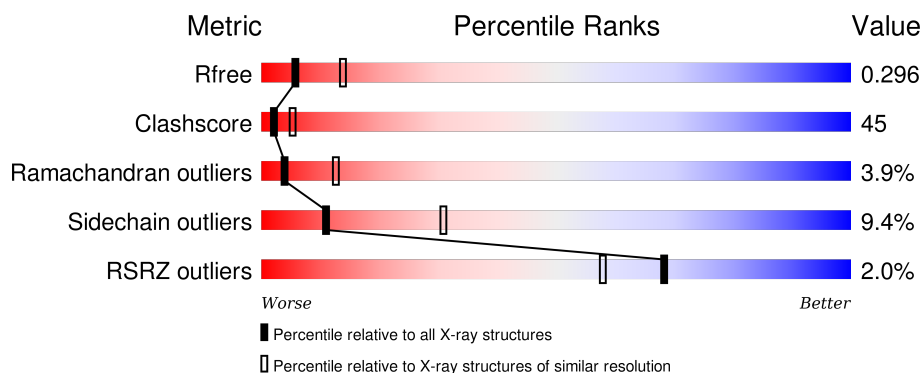
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.80 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}	91344	2393 (2.80-2.80)
Clashscore	102246	2827 (2.80-2.80)
Ramachandran outliers	100387	2782 (2.80-2.80)
Sidechain outliers	100360	2784 (2.80-2.80)
RSRZ outliers	91569	2404 (2.80-2.80)

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. 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	<div> <div></div> <div>43%</div> <div>47%</div> <div>9%</div> <div>.</div> </div>
2	B	428	<div> <div>3%</div> <div>39%</div> <div>45%</div> <div>9%</div> <div>6%</div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	NNC	A	1551	-	-	X	-

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 7872 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.

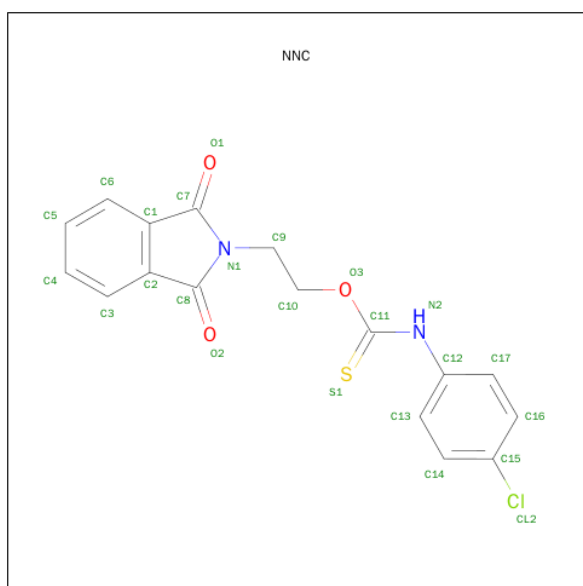
- Molecule 1 is a protein called REVERSE TRANSCRIPTASE/RIBONUCLEASE H.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	551	Total	C	N	O	S	0	0	0
			4475	2898	744	825	8			

- Molecule 2 is a protein called P51 RT.

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

- Molecule 3 is O-[2-(1,3-DIOXO-1,3-DIHYDRO-2H-ISOINDOL-2-YL)ETHYL] (4-CHLORO PHENYL)THIOCARBAMATE (three-letter code: NNC) (formula: C₁₇H₁₃ClN₂O₃S).



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

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	30	Total 30	O 30	0	0
4	B	25	Total 25	O 25	0	0

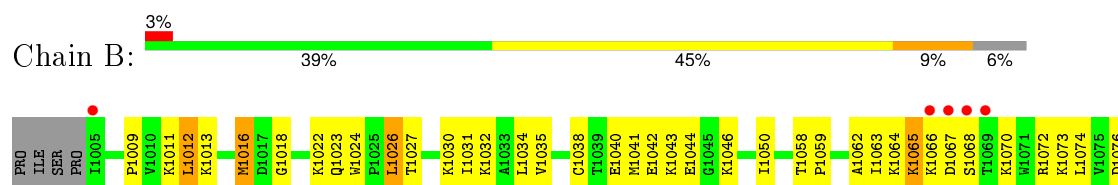
3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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



D1364	T1296	Y1232	K1154	F1077
V1365	E1297	E1233	G1155	R1078
K1366	E1298	L1234	E1079	L1080
Q1367	A1299	H1235	P1157	N1081
L1368	E1300	P1236	A1158	K1082
T1369	L1301	D1237	I1159	R1083
E1370	E1302	K1238	F1160	T1084
A1371	L1303	W1239	Q1161	Q1085
V1372	A1304	T1240	K1166	D1086
I1375	E1305	V1241	I1167	F1087
I1380	N1306	Q1242	L1168	W1088
W1381	R1307	P1243	E1169	E1089
I1382	E1308	I1244	P1170	V1090
K1385	I1309	V1245	N1175	Q1091
K1388	L1310	L1246	I1178	L1092
F1389	K1311	P1247	V1179	G1093
K1390	E1312	E1248	I1180	P1095
L1391	P1313	D1250	I1183	H1096
P1392	V1314	T1253	Y1184	K1101
I1393	V1317	V1254	D1185	K1104
Q1394	Y1318	M1255	L1187	S1105
K1395	K1323	D1256	Y1188	V1106
E1396	D1324	I1257	L1189	T1107
L1397	L1325	Q1258	G1205	V1108
W1398	I1326	K1259	R1206	L1109
W1401	Q1330	V1261	Q1207	D1110
W1402	K1331	G1262	S1191	V1111
E1403	Q1332	K1263	I1195	G1112
Y1404	G1333	L1264	I1202	D1113
Y1405	Q1334	M1265	L1205	A1114
W1406	G1335	W1266	L1206	Y1115
Q1407	Q1336	A1267	S1268	F1116
E1413	W1337	G1269	R1207	S1117
W1414	T1338	I1270	Q1207	V1118
M1418	Y1342	Y1271	R1211	E1122
T1419	Q1343	I1274	W1212	R1125
P1420	E1344	K1275	G1213	K1126
P1421	P1345	V1276	L1214	
L1422	F1346	R1277	THR	F1130
Y1423	K1347	Q1278	THR	T1131
K1424	M1348	L1279	PRO	I1132
L1425	L1349	C1280	ASP	
W1426	K1350	K1281	LYS	N1136
Y1427	T1351	L1282	LYS	N1137
Q1428	Y1354	LEU	HIS	
	A1355	ARG	GLN	Y1144
	ARG	GLY	LYS	Q1145
	MET	T1286	GLU	Y1146
	ARG	K1287	PRO	
	GLY	A1288	PRO	L1149
	GLY	L1289	PHE	F1150
	ALA		LEU	Q1151
	H1361	I1293	TRP	G1152
	T1362	P1294	M1230	W1153
	M1363	L1295	G1231	

4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α , β , γ	118.86Å 156.13Å 154.57Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.80 19.95 – 2.80	Depositor EDS
% Data completeness (in resolution range)	89.0 (20.00-2.80) 89.0 (19.95-2.80)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.60 (at 2.79Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.237 , 0.297 0.235 , 0.296	Depositor DCC
R_{free} test set	1589 reflections (5.28%)	DCC
Wilson B-factor (Å ²)	52.5	Xtriage
Anisotropy	0.004	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 46.6	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 31696 reflections	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	7872	wwPDB-VP
Average B, all atoms (Å ²)	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.83% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: NNC

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	0.26	1/4592 (0.0%)	0.40	0/6240
2	B	0.24	0/3411	0.40	0/4632
All	All	0.25	1/8003 (0.0%)	0.40	0/10872

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	478	GLU	CD-OE2	6.97	1.33	1.25

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	4475	0	4521	428	0
2	B	3318	0	3341	292	0
3	A	24	0	13	12	0
4	A	30	0	0	9	0
4	B	25	0	0	6	0
All	All	7872	0	7875	705	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 45.

The worst 5 of 705 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:107:THR:HG21	1:A:202:ILE:CD1	1.42	1.44
1:A:450:THR:CG2	1:A:452:LEU:HD22	1.51	1.38
1:A:357:MET:H	1:A:357:MET:CE	1.51	1.24
1:A:450:THR:HG21	1:A:452:LEU:HD22	1.22	1.18
2:B:1428:GLN:O	2:B:1428:GLN:NE2	1.77	1.17

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	547/557 (98%)	480 (88%)	48 (9%)	19 (4%)	4	15
2	B	393/428 (92%)	331 (84%)	44 (11%)	18 (5%)	3	9
All	All	940/985 (95%)	811 (86%)	92 (10%)	37 (4%)	4	12

5 of 37 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	84	THR
1	A	140	PRO
1	A	222	GLN
2	B	1065	LYS
2	B	1085	GLN

5.3.2 Protein sidechains ⓘ

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	489/497 (98%)	445 (91%)	44 (9%)	12	34
2	B	366/390 (94%)	330 (90%)	36 (10%)	10	28
All	All	855/887 (96%)	775 (91%)	80 (9%)	11	31

5 of 80 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	471	ASN
2	B	1011	LYS
2	B	1336	GLN
1	A	482	ILE
1	A	516	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 33 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	418	ASN
1	A	507	GLN
2	B	1340	GLN
1	A	428	GLN
1	A	447	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates

There are no carbohydrates in this entry.

5.6 Ligand geometry

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	NNC	A	1551	-	26,26,26	2.40	6 (23%)	36,36,36	3.06	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. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	NNC	A	1551	-	-	0/9/26/26	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	NNC	C1-C7	-6.01	1.39	1.48
3	A	1551	NNC	C2-C8	-5.89	1.39	1.48
3	A	1551	NNC	C8-N1	-5.02	1.33	1.39
3	A	1551	NNC	C7-N1	-4.75	1.33	1.39
3	A	1551	NNC	C12-N2	-4.34	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	NNC	C10-O3-C11	-11.30	109.55	119.09
3	A	1551	NNC	O3-C11-S1	-7.30	119.88	125.06
3	A	1551	NNC	C12-N2-C11	-5.84	120.73	130.04
3	A	1551	NNC	C8-N1-C7	-3.44	109.07	111.95

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	1551	NNC	C10-C9-N1	-2.70	108.71	112.37

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 12 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	1551	NNC	12	0

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(Å ²)	Q<0.9
1	A	551/557 (98%)	-0.08	8 (1%) 76 68	20, 44, 67, 85	0
2	B	401/428 (93%)	-0.16	11 (2%) 58 45	23, 40, 75, 90	0
All	All	952/985 (96%)	-0.11	19 (1%) 68 58	20, 43, 72, 90	0

The worst 5 of 19 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	286	THR	6.1
1	A	24	TRP	4.6
2	B	1067	ASP	3.8
2	B	1230	MET	3.7
1	A	472	THR	3.6

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(\AA^2)	Q<0.9
3	NNC	A	1551	24/24	0.97	0.14	-0.93	26,29,36,38	0

6.5 Other polymers ⓘ

There are no such residues in this entry.