



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 27, 2018 – 03:19 AM EST

PDB ID : 1S9G  
Title : CRYSTAL STRUCTURE OF HIV-1 REVERSE TRANSCRIPTASE (RT) IN COMPLEX WITH JANSSEN-R120394.  
Authors : Das, K.; Clark Jr., A.D.; Ludovici, D.W.; Kukla, M.J.; Decorte, B.; Lewi, P.J.; Hughes, S.H.; Janssen, P.A.; Arnold, E.  
Deposited on : 2004-02-04  
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](mailto: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	:	1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030736
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	rb-20030736

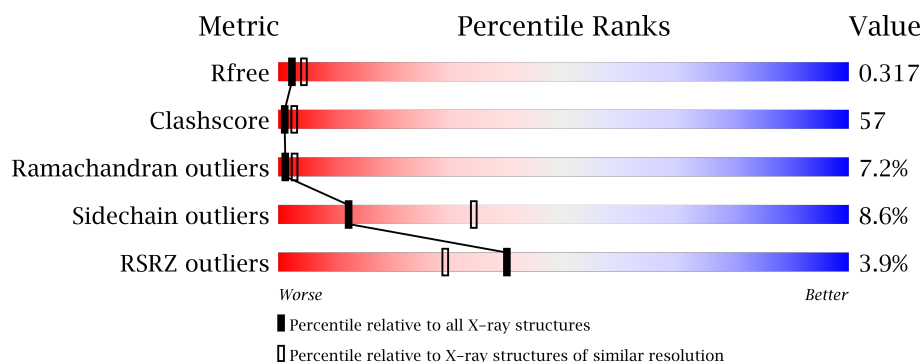
# 1 Overall quality at a glance ⓘ

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))
$R_{free}$	100719	2583 (2.80-2.80)
Clashscore	112137	3033 (2.80-2.80)
Ramachandran outliers	110173	2983 (2.80-2.80)
Sidechain outliers	110143	2985 (2.80-2.80)
RSRZ outliers	101464	2610 (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  $\geq 3$ , 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	560	<div> <div>3%</div> <div>31%</div> <div>54%</div> <div>12%</div> <div>.</div> </div>
2	B	430	<div> <div>5%</div> <div>35%</div> <div>52%</div> <div>11%</div> <div>..</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	ABZ	A	701	-	-	X	-

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 8112 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 POL polyprotein [Contains: Reverse transcriptase].

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	543	Total	C	N	O	S	66	0	0
			4429	2871	736	815	7			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	280	SER	CYS	ENGINEERED	UNP P03366

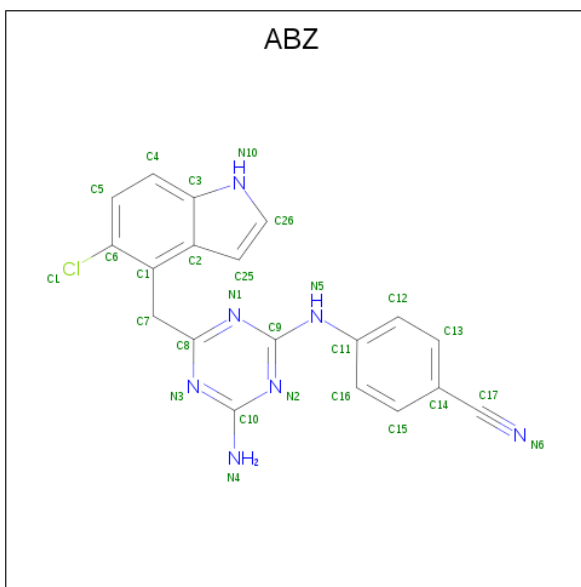
- Molecule 2 is a protein called POL polyprotein [Contains: Reverse transcriptase].

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	425	Total	C	N	O	S	55	0	0
			3514	2289	582	636	7			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	280	SER	CYS	ENGINEERED	UNP P03366

- Molecule 3 is 4-[4-AMINO-6-(5-CHLORO-1H-INDOL-4-YLMETHYL)-[1,3,5]TRIAZIN-2-YLAMINO]-BENZONITRILE (three-letter code: ABZ) (formula: C<sub>19</sub>H<sub>14</sub>ClN<sub>7</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	Cl	N	0	0
			27	19	1	7		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	79	Total	O	0	0
			79	79		
4	B	63	Total	O	0	0
			63	63		



T419	P420	L422	V423	K424	L425	W426	Y427	GLN	LEU	GLU	L349	K350	T351	G352	K353	Y354	A355	K356	M357	K358	G359	G359	A360	K361	T362	K363	D364	V365	K366	Q367	L368	T369	E370	A371	V372	Q373	K374	T375	T376	T377	E378	S379	W383	G384	K385	T386	P387	K388	F389	L391	P392	I393	Q394	W398	E399	T400	W401	E404	Y405	W406	W410	T411	P412	E413	W418	L283	K284	A288	L289	T290	E291	V292	L293	P294	L295	T296	E297	E298	A299	E300	L301	E302	L303	K304	E305	K306	K307	E308	L309	L310	K311	V314	K315	G316	V317	T318	Y319	D320	P321	S322	K323	D324	L325	L329	Q330	K331	Q332	G335	Q336	K337	T338	Y339	Q340	L341	Q342	Q343	E344	P345	F346	K347	N348	K220	H221	K222	K223	E224	P225	P226	F227	L228	W229	K230	G231	Y232	E233	P236	D237	K238	W239	T240	V241	Q242	P243	I244	V245	L246	P247	E248	K249	T253	V254	N255	D256	I257	Q258	K259	L260	V261	G262	K263	L264	N265	W266	A267	S268	Q269	I270	Y271	P272	G273	I274	K275	V276	R277	Q278	L279	S280	K281	L282	P150	W153	K154	G155	S156	P157	A158	I159	F160	Q161	S162	S163	L167	L168	E169	P170	K171	K172	K173	Q174	N175	P176	D177	I178	V179	L180	M184	D185	D186	L187	Y188	S191	D192	L193	E194	I195	G196	Q197	L198	H198	R199	T200	K201	I202	E203	E204	L205	K206	Q207	H208	L209	I214	P217	D218	K219	V76	F77	R78	T84	D85	D86	F87	W88	E89	V90	Q91	L92	H96	P97	L100	K101	K102	K103	K104	S105	V106	T107	V108	L109	D110	V111	G112	D113	A114	Y115	F116	L120	R125	K126	Y127	T128	A129	F130	T131	S134	I135	N136	N137	E138	T139	P140	G141	I142	R143	Y144	Q145	Y146	N147
------	------	------	------	------	------	------	------	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	223.60Å 68.60Å 103.00Å 90.00° 107.50° 90.00°	Depositor
Resolution (Å)	20.00 – 2.80 25.21 – 2.80	Depositor EDS
% Data completeness (in resolution range)	88.8 (20.00-2.80) 91.9 (25.21-2.80)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.25 (at 2.80Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, $R_{free}$	0.242 , 0.312 0.250 , 0.317	Depositor DCC
$R_{free}$ test set	1723 reflections (5.08%)	DCC
Wilson B-factor (Å <sup>2</sup> )	64.1	Xtriage
Anisotropy	0.114	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 69.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	8112	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	74.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.72% 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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ABZ

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	0.53	0/4547	0.83	4/6178 (0.1%)
2	B	0.64	0/3618	0.90	5/4918 (0.1%)
All	All	0.58	0/8165	0.86	9/11096 (0.1%)

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
2	B	0	1

There are no bond length outliers.

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	54	ASN	N-CA-C	-7.76	90.06	111.00
1	A	359	GLY	N-CA-C	-7.62	94.05	113.10
1	A	286	THR	N-CA-C	-7.56	90.60	111.00
2	B	89	GLU	N-CA-C	-6.43	93.64	111.00
2	B	230	MET	N-CA-C	6.23	127.82	111.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	130	PHE	Sidechain

## 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	4429	0	4491	524	0
2	B	3514	0	3547	403	0
3	A	27	0	14	11	0
4	A	79	0	0	3	0
4	B	63	0	0	2	0
All	All	8112	0	8052	893	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 57.

The worst 5 of 893 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:297:GLU:HG2	2:B:298:GLU:N	1.56	1.13
1:A:2:ILE:HG21	1:A:5:ILE:HG13	1.31	1.12
2:B:206:ARG:HB3	2:B:206:ARG:NH1	1.64	1.11
2:B:297:GLU:HG2	2:B:298:GLU:H	0.95	1.10
1:A:311:LYS:HA	1:A:311:LYS:CE	1.82	1.09

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	541/560 (97%)	415 (77%)	89 (16%)	37 (7%)	<b>1</b> <b>3</b>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	B	423/430 (98%)	323 (76%)	68 (16%)	32 (8%)	1	2
All	All	964/990 (97%)	738 (77%)	157 (16%)	69 (7%)	1	3

5 of 69 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	14	PRO
1	A	55	PRO
1	A	113	ASP
1	A	141	GLY
1	A	151	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	486/500 (97%)	444 (91%)	42 (9%)	12	34
2	B	387/392 (99%)	354 (92%)	33 (8%)	12	35
All	All	873/892 (98%)	798 (91%)	75 (9%)	12	34

5 of 75 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	451	LYS
1	A	529	GLU
2	B	374	LYS
1	A	460	ASN
1	A	510	PRO

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 23 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	136	ASN
2	B	207	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
2	B	407	GLN
2	B	174	GLN
2	B	208	HIS

### 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	ABZ	A	701	-	27,30,30	1.91	10 (37%)	34,42,42	2.61	14 (41%)

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	ABZ	A	701	-	-	0/10/10/10	0/4/4/4

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	701	ABZ	C9-N5	2.16	1.40	1.36
3	A	701	ABZ	C4-C5	2.18	1.41	1.36
3	A	701	ABZ	C12-C11	2.18	1.42	1.39
3	A	701	ABZ	C13-C12	2.24	1.42	1.38
3	A	701	ABZ	C13-C14	2.28	1.44	1.39

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	701	ABZ	N2-C9-N1	-4.17	119.92	126.23
3	A	701	ABZ	C5-C4-C3	-4.02	115.91	120.84
3	A	701	ABZ	N3-C10-N2	-3.93	119.08	125.45
3	A	701	ABZ	C11-N5-C9	-3.78	119.33	129.17
3	A	701	ABZ	N3-C8-N1	-3.59	122.08	126.35

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 11 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	701	ABZ	11	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, 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	543/560 (96%)	-0.17	15 (2%)	53 43	29, 78, 110, 112	17 (3%)
2	B	425/430 (98%)	-0.08	23 (5%)	26 17	16, 63, 111, 112	14 (3%)
All	All	968/990 (97%)	-0.13	38 (3%)	40 29	16, 75, 111, 112	31 (3%)

The worst 5 of 38 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	3	SER	5.4
2	B	229	TRP	5.1
1	A	140	PRO	4.2
2	B	225	PRO	4.2
1	A	2	ILE	3.9

### 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, 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	ABZ	A	701	27/27	0.75	0.34	1.33	43,63,70,73	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.