



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 04:30 pm GMT

PDB ID : 3LAK  
Title : Crystal structure of HIV-1 reverse transcriptase in complex with N1-heterocycle pyrimidinedione non-nucleoside inhibitor  
Authors : Lansdon, E.B.; Mitchell, M.L.  
Deposited on : 2010-01-06  
Resolution : 2.30 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 : trunk28620  
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) : recalc28949



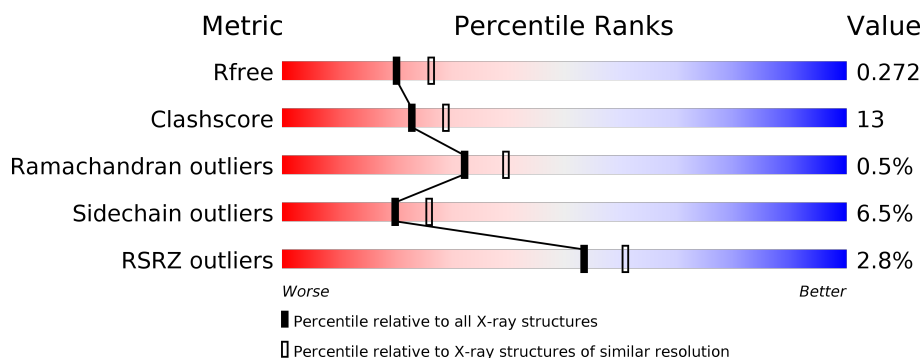
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.30 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	4130 (2.30-2.30)
Clashscore	112137	4751 (2.30-2.30)
Ramachandran outliers	110173	4705 (2.30-2.30)
Sidechain outliers	110143	4704 (2.30-2.30)
RSRZ outliers	101464	4156 (2.30-2.30)

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>2%</div> <div>67%</div> <div>28%</div> <div>••</div> </div>
1	B	560	<div> <div>3%</div> <div>53%</div> <div>18%</div> <div>•</div> <div>28%</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	SO4	A	563	-	-	-	X



## 2 Entry composition [i](#)

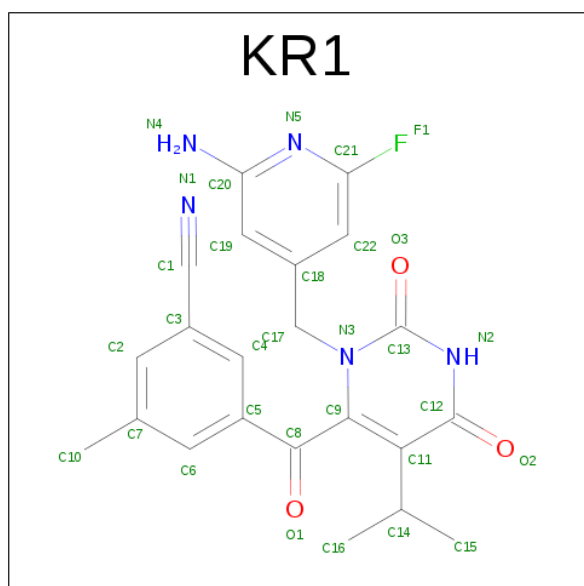
There are 5 unique types of molecules in this entry. The entry contains 8101 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 HIV Reverse transcriptase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	553	Total	C	N	O	S	0	0	0
			4508	2914	752	834	8			
1	B	403	Total	C	N	O	S	0	0	0
			3334	2173	550	605	6			

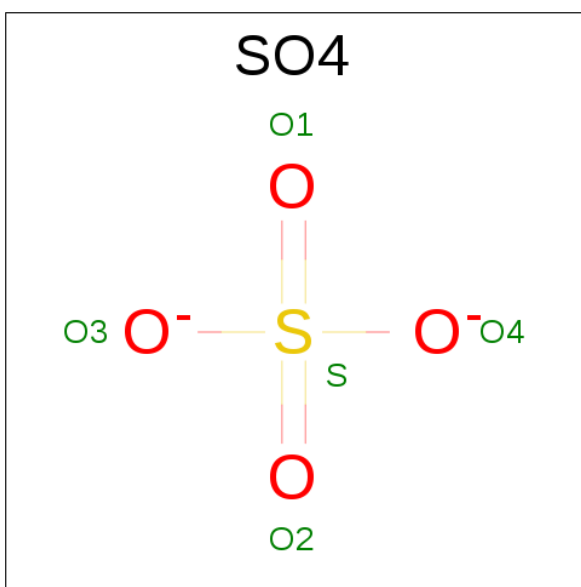
- Molecule 2 is 3-({3-[(2-AMINO-6-FLUOROPYRIDIN-4-YL)METHYL]-5-(1-METHYLETHYL)-2,6-DIOXO-1,2,3,6-TETRAHYDOPYRIMIDIN-4-YL}CARBONYL)-5-METHYLBENZONITRILE (three-letter code: KR1) (formula: C<sub>22</sub>H<sub>20</sub>FN<sub>5</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	F	N	O	0	0
			31	22	1	5	3		

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).





Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	O	S	0	0
			5	4	1		
3	A	1	Total	O	S	0	0
			5	4	1		
3	A	1	Total	O	S	0	0
			5	4	1		
3	A	1	Total	O	S	0	0
			5	4	1		

- Molecule 4 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Cl	0	0
			1	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	110	Total	O	0	0
			110	110		
5	B	97	Total	O	0	0
			97	97		



These plots are drawn for all protein, RNA and DNA 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 ( $\text{RSRZ} > 2$ ). 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.

[illegible]

Chain B:

53% 18% 28%

PRO ILE SER PRO I5 E6 T7 K11 L12 K13 P14 W24 K30 A33 I37 T58 P59 V60 K65 LYS ASP S68 T69 K70 V75 L80 R83 T84 Q85 W88 E89 Y90 Q91 K101 S105 V106 Y115 F116 P119 R143 L149 P150 Q151

S156 P157 I167 L168 E169 P170 M175 Y183 M184 L187 V188 Y189 G190 E194 Q197 K201 T202 E203 E204 L205 L209 W212 G213 L214 T215 THR PRO ASP LYS LYS HIS GLN LYS GLU PRO PRO PHE LEU TRP MET GLY Y232 E233 L234 H235 P236 V241 T242

T244 V245 K249 D250 S251 W252 T253 V254 Q258 K259 L260 V261 K275 V276 K277 Q278 L279 C280 K281 L282 L283 R284 Q285 T286 K287 A288 T293 P294 L295 T296 E297 E298 A299 E300 L301 E305 N306 K307 E308 L309 L310 K311 E312 P313 K323 Q330 K331 Q332 G335 Q336 W337 T338 E344 P345 F346 K347 K353 Y354 A355 R356 MET ARG GLY ALA H361 Q367 L368 T369 Q373 T376 K385 T386 P387 K388 K395 E399 T403 F404 Y405 W406 Q407 E415 L429 LYS GLU GLU PRO PRO ILE VAL GLY ALA GLU THR PHE TYR VAL ASP GLY ALA SER CYS



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



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	118.03Å 154.60Å 154.48Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.91 – 2.30 47.23 – 2.28	Depositor EDS
% Data completeness (in resolution range)	90.4 (29.91-2.30) 89.2 (47.23-2.28)	Depositor EDS
$R_{merge}$	0.04	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.81 (at 2.27Å)	Xtriage
Refinement program	CNX 2005	Depositor
R, $R_{free}$	0.223 , 0.274 0.221 , 0.272	Depositor DCC
$R_{free}$ test set	2875 reflections (5.33%)	DCC
Wilson B-factor (Å <sup>2</sup> )	44.7	Xtriage
Anisotropy	0.355	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 43.1	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	0.94	EDS
Total number of atoms	8101	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	51.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.48% 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: KR1, SO4, CL

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	0.33	0/4625	0.49	0/6284
1	B	0.34	0/3426	0.49	0/4655
All	All	0.33	0/8051	0.49	0/10939

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	B	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	361	HIS	Peptide

### 5.2 Too-close contacts

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	4508	0	4558	133	0
1	B	3334	0	3365	75	0
2	A	31	0	20	0	0
3	A	20	0	0	0	0
4	B	1	0	0	0	0
5	A	110	0	0	3	0
5	B	97	0	0	4	0
All	All	8101	0	7943	204	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 13.

The worst 5 of 204 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:131:THR:HG22	1:A:143:ARG:HG2	1.23	1.08
1:A:66:LYS:HG3	1:A:68:SER:H	1.40	0.87
1:B:296:THR:HG22	1:B:298:GLU:H	1.39	0.86
1:B:209:LEU:HD12	1:B:214:LEU:HD22	1.56	0.85
1:B:194:GLU:HG3	1:B:197:GLN:HG3	1.60	0.84

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	551/560 (98%)	527 (96%)	19 (3%)	5 (1%)	20	23
1	B	395/560 (70%)	381 (96%)	14 (4%)	0	100	100
All	All	946/1120 (84%)	908 (96%)	33 (4%)	5 (0%)	32	39

All (5) Ramachandran outliers are listed below:



Mol	Chain	Res	Type
1	A	91	GLN
1	A	356	ARG
1	A	543	GLY
1	A	66	LYS
1	A	465	LYS

### 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	495/500 (99%)	457 (92%)	38 (8%)	15	18
1	B	367/500 (73%)	349 (95%)	18 (5%)	29	39
All	All	862/1000 (86%)	806 (94%)	56 (6%)	20	26

5 of 56 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	399	GLU
1	A	500	GLN
1	B	310	LEU
1	A	410	TRP
1	A	448	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 24 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	85	GLN
1	B	175	ASN
1	B	367	GLN
1	B	91	GLN
1	B	151	GLN

### 5.3.3 RNA ⓘ

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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 6 ligands modelled in this entry, 1 is monoatomic - leaving 5 for Mogul analysis.

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	SO4	A	561	-	4,4,4	0.32	0	6,6,6	0.17	0
3	SO4	A	562	-	4,4,4	0.33	0	6,6,6	0.15	0
3	SO4	A	563	-	4,4,4	0.36	0	6,6,6	0.08	0
3	SO4	A	564	-	4,4,4	0.36	0	6,6,6	0.08	0
2	KR1	A	701	-	29,33,33	2.84	13 (44%)	37,48,48	3.19	11 (29%)

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	SO4	A	561	-	-	0/0/0/0	0/0/0/0
3	SO4	A	562	-	-	0/0/0/0	0/0/0/0
3	SO4	A	563	-	-	0/0/0/0	0/0/0/0
3	SO4	A	564	-	-	0/0/0/0	0/0/0/0
2	KR1	A	701	-	-	0/18/18/18	0/3/3/3

The worst 5 of 13 bond length outliers are listed below:



Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	701	KR1	C9-C8	-2.54	1.46	1.50
2	A	701	KR1	C2-C7	2.32	1.43	1.39
2	A	701	KR1	C15-C14	2.44	1.59	1.52
2	A	701	KR1	F1-C21	2.77	1.40	1.35
2	A	701	KR1	C1-N1	2.93	1.21	1.14

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	701	KR1	O1-C8-C9	-6.39	112.53	120.12
2	A	701	KR1	C22-C21-N5	-4.82	120.62	127.11
2	A	701	KR1	C17-N3-C13	-4.12	113.10	117.92
2	A	701	KR1	C21-C22-C18	-3.75	117.02	119.32
2	A	701	KR1	C5-C4-C3	-2.21	117.91	120.42

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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	553/560 (98%)	0.08	10 (1%) 69 74	30, 49, 82, 101	0
1	B	403/560 (71%)	0.16	17 (4%) 37 44	26, 47, 82, 95	0
All	All	956/1120 (85%)	0.12	27 (2%) 53 61	26, 48, 82, 101	0

The worst 5 of 27 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	14	PRO	5.7
1	B	232	TYR	4.4
1	B	279	LEU	3.8
1	A	452	LEU	3.8
1	B	13	LYS	3.7

### 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	SO4	A	563	5/5	0.83	0.31	5.05	120,120,120,120	0
3	SO4	A	561	5/5	0.96	0.17	1.13	61,65,67,68	0
2	KR1	A	701	31/31	0.94	0.15	0.45	34,41,44,48	0
3	SO4	A	562	5/5	0.96	0.12	-0.39	58,60,61,64	0
4	CL	B	561	1/1	0.95	0.24	-	62,62,62,62	0
3	SO4	A	564	5/5	0.73	0.16	-	107,107,108,108	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.