wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 31, 2016 – 07:52 PM GMT

PDB ID	:	1HNI
Title	:	STRUCTURE OF HIV-1 REVERSE TRANSCRIPTASE IN A COMPLEX WITH THE NONNUCLEOSIDE INHIBITOR ALPHA-APA R 95845 AT 2.8 ANGSTROMS RESOLUTION
Authors	:	Ding, J.; Das, K.; Arnold, E.
Deposited on	:	1995-02-28
Resolution	:	2.80 Å(reported)

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 (i)

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.
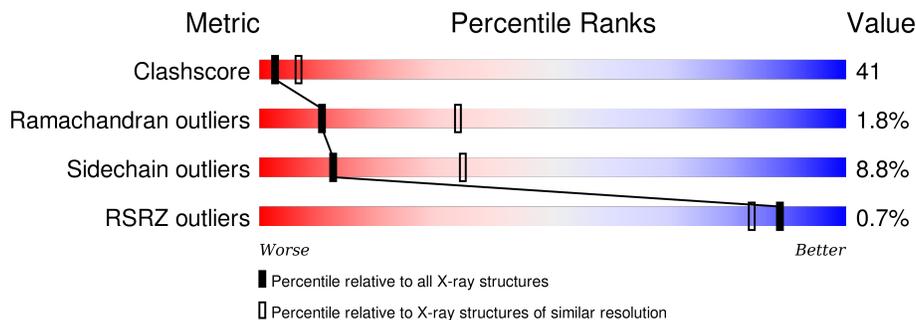
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	102246	2827 (2.80-2.80)
Ramachandran outliers	100387	2782 (2.80-2.80)
Sidechain outliers	100360	2784 (2.80-2.80)
RSRZ outliers	91569	2404 (2.80-2.80)

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 <=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	558	39% 53% 7%
2	B	427	41% 54% 5%

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	AAA	A	559	-	-	X	-

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 7819 atoms, of which 4 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-1 REVERSE TRANSCRIPTASE (SUBUNIT P66).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	558	Total	C	N	O	S	0	0	0
			4368	2825	724	813	6			

There is a discrepancy between the modelled and reference sequences:

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

- Molecule 2 is a protein called HIV-1 REVERSE TRANSCRIPTASE (SUBUNIT P51).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	427	Total	C	N	O	S	0	0	0
			3424	2231	564	624	5			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	172	ARG	LYS	CONFLICT	UNP P03367
B	280	SER	CYS	CONFLICT	UNP P03367
B	357	THR	MET	CONFLICT	UNP P03367

- Molecule 3 is (2-ACETYL-5-METHYLANILINO)(2,6-DIBROMOPHENYL)ACETAMIDE (three-letter code: AAA) (formula: $C_{17}H_{16}Br_2N_2O_2$).
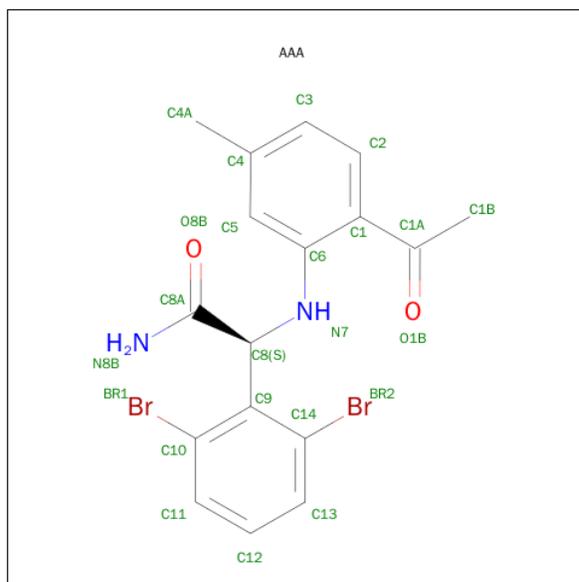
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
3	A	1	Total	Br	C	H	N	O	0	0
			27	2	17	4	2	2		

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: HIV-1 REVERSE TRANSCRIPTASE (SUBUNIT P66)
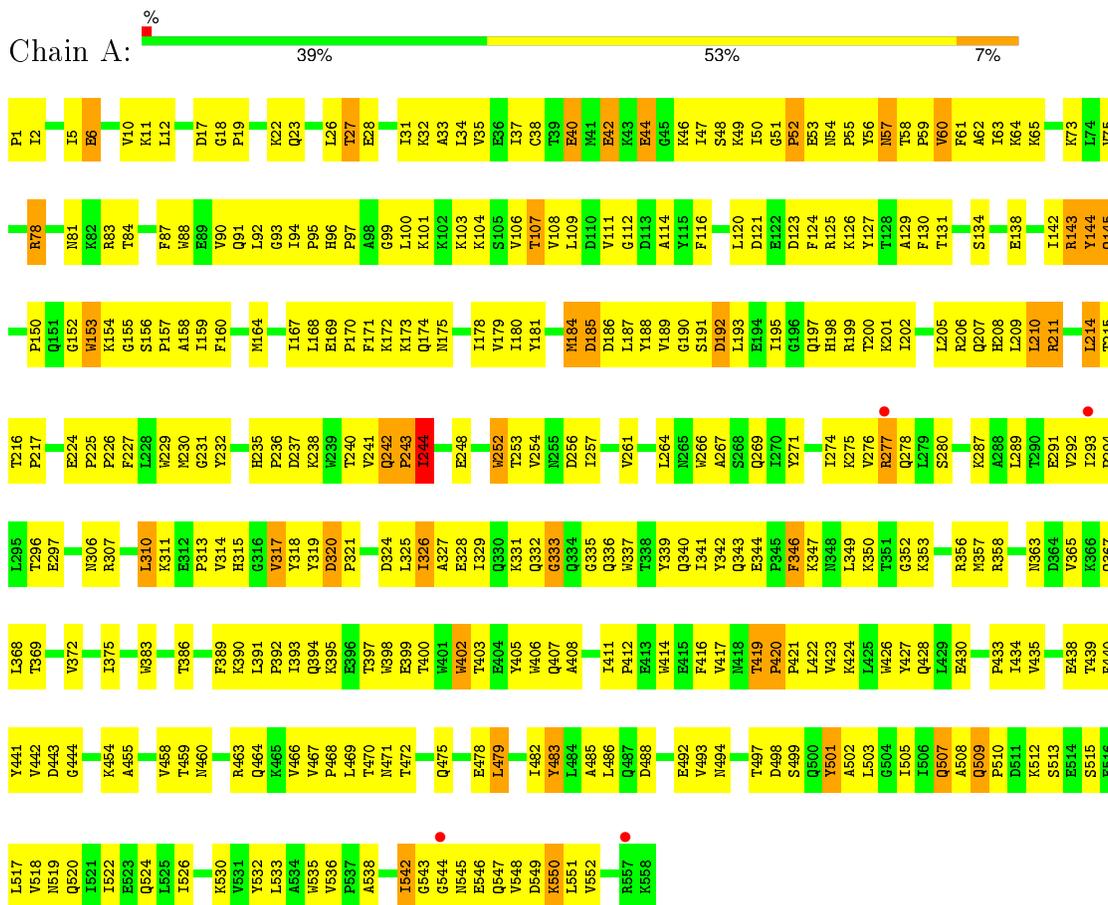
- Molecule 2: HIV-1 REVERSE TRANSCRIPTASE (SUBUNIT P51)

P150 Q151 G152 W153 K154 G155 S156 P157 A158 I159 F160 Q161 S162 S163 M164 T165 K166 I167 L168 E169 P170 F171 R172 K173 Q174 N175 P176 D177 I178 V179 I180 Y181 Q182 Y183 M184 D185 D186 L187 Y188 V189 G190 S191 D192 L193 E194 Q197 T200 K201 I202 E203 E204 L205 R206 Q207 H208 L209 L210 R211

L214 T215 T216 P217 K220 H221 Q222 K223 E224 P225 P226 N230 G231 Y232 E233 P236 W239 T240 V241 Q242 P243 I244 V245 L246 D250 S251 W252 T253 V254 N255 D256 I257 Q258 K259 L260 V261 G262 K263 L264 N265 I274 K275 V276 R277 Q278 L279 S280 K281 L282 L283 R284 L289

T290 E291 V292 I293 P294 L295 T296 E297 E298 A299 E300 L301 E302 L303 N306 R307 E308 I309 L310 V314 H315 G316 V317 Y318 Y319 D320 P321 S322 K323 D324 L325 I329 Q330 K331 Q332 Q336 W337 T338 Y339 Q340 I341 Y342 Q343 E344 P345 F346 K347 N348 L349 K350 T351 G352 K353 Y354 A355 R356

D364 V365 K366 Q367 L368 T369 E370 A371 V372 Q373 T376 S379 I380 W383 G384 I393 Q394 K395 E396 T397 W398 E399 W402 T403 E404 Y405 W406 Q407 A408 I411 P412 E413 W414 N418 T419 P420 K424 L425 W426 Y427

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	223.30Å 69.90Å 106.50Å 90.00° 105.40° 90.00°	Depositor
Resolution (Å)	(Not available) – 2.80 36.66 – 2.60	Depositor EDS
% Data completeness (in resolution range)	(Not available) ((Not available)-2.80) 75.7 (36.66-2.60)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	3.91 (at 2.61Å)	Xtriage
Refinement program	X-PLOR	Depositor
R, R_{free}	0.255 , (Not available) 0.284 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	DCC
Wilson B-factor (Å^2)	30.3	Xtriage
Anisotropy	0.172	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 595.1	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$<\lvert L \rvert> = 0.40, <L^2> = 0.22$	Xtriage
Outliers	0 of 37099 reflections	Xtriage
F_o,F_c correlation	0.80	EDS
Total number of atoms	7819	wwPDB-VP
Average B, all atoms (Å^2)	39.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.93% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $<\lvert L \rvert>$, $<L^2>$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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:
AAA

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	0.67	0/4483	0.50	1/6115 (0.0%)
2	B	1.41	4/3523 (0.1%)	1.25	8/4800 (0.2%)
All	All	1.06	4/8006 (0.0%)	0.91	9/10915 (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
1	A	0	2
2	B	0	1
All	All	0	3

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	172	ARG	CD-NE	60.74	2.49	1.46
2	B	22	LYS	CE-NZ	28.75	2.21	1.49
2	B	172	ARG	NE-CZ	12.94	1.49	1.33
2	B	22	LYS	CD-CE	-10.15	1.25	1.51

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	172	ARG	CD-NE-CZ	-63.77	34.32	123.60
2	B	22	LYS	CD-CE-NZ	-36.13	28.59	111.70
2	B	172	ARG	CG-CD-NE	15.56	144.48	111.80
2	B	78	ARG	NE-CZ-NH2	7.46	124.03	120.30
2	B	83	ARG	NE-CZ-NH2	7.32	123.96	120.30

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	483	TYR	Sidechain
1	A	501	TYR	Sidechain
2	B	127	TYR	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	4368	0	4256	388	0
2	B	3424	0	3372	267	0
3	A	23	4	16	12	0
All	All	7815	4	7644	631	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 41.

The worst 5 of 631 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:85:GLN:HG3	2:B:154:LYS:CB	1.67	1.25
1:A:420:PRO:HB2	1:A:421:PRO:HD3	1.27	1.11
1:A:112:GLY:HA2	1:A:185:ASP:OD2	1.56	1.04
2:B:59:PRO:HG2	2:B:76:ASP:HB3	1.35	1.02
1:A:536:VAL:HG21	1:A:542:ILE:HG12	1.41	1.01

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	556/558 (100%)	443 (80%)	99 (18%)	14 (2%)	7	24
2	B	425/427 (100%)	354 (83%)	67 (16%)	4 (1%)	21	55
All	All	981/985 (100%)	797 (81%)	166 (17%)	18 (2%)	11	34

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	184	MET
1	A	53	GLU
1	A	153	TRP
1	A	242	GLN
1	A	287	LYS

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	453/498 (91%)	409 (90%)	44 (10%)	10	29
2	B	362/389 (93%)	334 (92%)	28 (8%)	16	41
All	All	815/887 (92%)	743 (91%)	72 (9%)	12	35

5 of 72 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	386	THR
1	A	509	GLN
2	B	340	GLN
1	A	402	TRP
1	A	472	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 21 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	507	GLN
1	A	520	GLN
2	B	315	HIS
1	A	474	ASN
2	B	343	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 carbohydrates 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	AAA	A	559	-	23,24,24	1.34	3 (13%)	29,34,34	1.29	3 (10%)

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	AAA	A	559	-	-	0/16/16/16	0/2/2/2

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	559	AAA	C1-C6	2.29	1.44	1.41
3	A	559	AAA	C6-N7	2.51	1.43	1.37
3	A	559	AAA	C5-C6	3.78	1.45	1.39

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	559	AAA	O8B-C8A-N8B	-2.00	120.11	123.08
3	A	559	AAA	O1B-C1A-C1	2.34	123.66	120.50
3	A	559	AAA	C8-C8A-N8B	3.10	122.08	116.62

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	559	AAA	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(Å²)	Q<0.9
1	A	558/558 (100%)	-0.43	4 (0%)	89	84	8, 44, 66, 72	0
2	B	427/427 (100%)	-0.47	3 (0%)	89	84	5, 33, 64, 72	0
All	All	985/985 (100%)	-0.44	7 (0%)	89	84	5, 41, 65, 72	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	231	GLY	3.8
2	B	282	LEU	3.7
1	A	293	ILE	3.2
1	A	544	GLY	3.0
1	A	557	ARG	2.9

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, 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(Å^2)	Q<0.9
3	AAA	A	559	23/23	0.94	0.15	-0.04	10,20,28,32	0

6.5 Other polymers (i)

There are no such residues in this entry.