



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 12:24 PM GMT

PDB ID : 3R8D  
Title : Activation of the Human Nuclear Xenobiotic Receptor PXR by the Reverse Transcriptase-Targeted Anti-HIV Drug PNU-142721  
Authors : Cheng, Y.; Redinbo, M.R.  
Deposited on : 2011-03-23  
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 (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

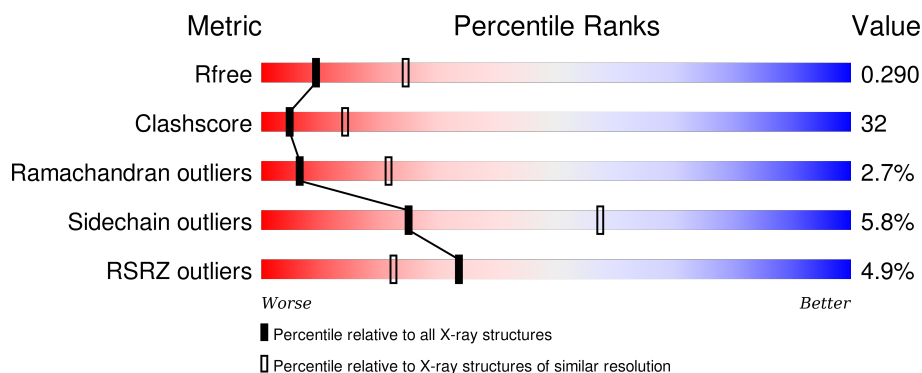
# 1 Overall quality at a glance ⓘ

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  $\geq 3$ , 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	316	

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
2	PNU	A	1	-	-	X	X

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 2140 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.

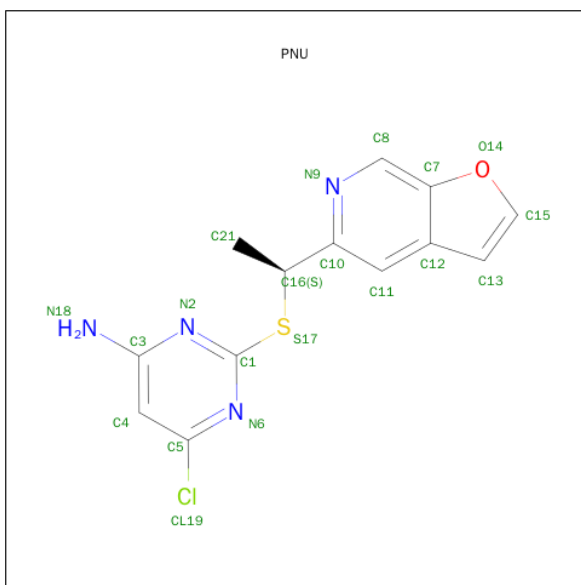
- Molecule 1 is a protein called Nuclear receptor subfamily 1 group I member 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	264	2105	1358	356	374	17	1	0	0

There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	119	MET	-	EXPRESSION TAG	UNP O75469
A	120	LYS	-	EXPRESSION TAG	UNP O75469
A	121	LYS	-	EXPRESSION TAG	UNP O75469
A	122	GLY	-	EXPRESSION TAG	UNP O75469
A	123	HIS	-	EXPRESSION TAG	UNP O75469
A	124	HIS	-	EXPRESSION TAG	UNP O75469
A	125	HIS	-	EXPRESSION TAG	UNP O75469
A	126	HIS	-	EXPRESSION TAG	UNP O75469
A	127	HIS	-	EXPRESSION TAG	UNP O75469
A	128	HIS	-	EXPRESSION TAG	UNP O75469
A	129	GLY	-	EXPRESSION TAG	UNP O75469

- Molecule 2 is 6-CHLORO-2-(1-FURO[2,3-C]PYRIDIN-5-YL-ETHYLSULFANYL)-PYRIMIDIN-4-YLAMINE (three-letter code: PNU) (formula: C<sub>13</sub>H<sub>11</sub>ClN<sub>4</sub>OS).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
2	A	1	Total	C	Cl	N	O	S	0	0
			20	13	1	4	1	1		

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	15	Total	O	0	0
			15	15		

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 ( $\text{RSRZ} > 2$ ). 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.

- Chain A:
- 
- 4% 43% 35% 16%
- LEU PRO GLU LYS SER LEU HIS GLN PRO LEU GLY VAL GIN G142 L143 T144 E145 E146 M149 M150 L151 R152 E153 L154 M155 D156 A157 K158 Q159 K160 T161 F162 D163 S167 K170 R173 L174 P175 G176 VAL LEU SER SER GLY CYS
- LEU PRO GLU LYS SER LEU HIS GLN PRO LEU GLY VAL GIN G142 L143 T144 E145 E146 M149 M150 L151 R152 E153 L154 M155 D156 A157 K158 Q159 K160 T161 F162 D163 S167 K170 R173 L174 P175 G176 VAL LEU SER SER GLY CYS
- M250 E337 E338 E339 Y340 V341 L342 M343 Q344 D352 R353 Q358 H359 R360 V361 V362 A370 Y376 C379 M380 Q383 R387 F388 L389 F390 L391 K392 L393 M394 A395 M396 R399 L400 L401 S402 L403 R410 L411 L412 L413 L414 F420 A421 T422 P423 L424 M425 F429 C430 V431
- LEU PRO GLU LYS SER LEU HIS GLN PRO LEU GLY VAL GIN G142 L143 T144 E145 E146 M149 M150 L151 R152 E153 L154 M155 D156 A157 K158 Q159 K160 T161 F162 D163 S167 K170 R173 L174 P175 G176 VAL LEU SER SER GLY CYS
- M250 E337 E338 E339 Y340 V341 L342 M343 Q344 D352 R353 Q358 H359 R360 V361 V362 A370 Y376 C379 M380 Q383 R387 F388 L389 F390 L391 K392 L393 M394 A395 M396 R399 L400 L401 S402 L403 R410 L411 L412 L413 L414 F420 A421 T422 P423 L424 M425 F429 C430 V431

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	92.28Å 92.28Å 84.38Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	41.27 – 2.80 41.27 – 2.80	Depositor EDS
% Data completeness (in resolution range)	99.7 (41.27-2.80) 99.7 (41.27-2.80)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	5.10 (at 2.81Å)	Xtriage
Refinement program	REFMAC 5.5.0072	Depositor
R, $R_{free}$	0.238 , 0.289 0.248 , 0.290	Depositor DCC
$R_{free}$ test set	450 reflections (5.03%)	DCC
Wilson B-factor (Å <sup>2</sup> )	55.7	Xtriage
Anisotropy	0.003	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 39.5	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 9394 reflections	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	2140	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	47.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.67% 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, 0.375 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: PNU

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	1.50	12/2149 (0.6%)	1.22	9/2894 (0.3%)

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	145	GLU	CD-OE1	5.96	1.32	1.25
1	A	263	TYR	CE2-CZ	5.91	1.46	1.38
1	A	289	ASN	CG-OD1	-5.89	1.10	1.24
1	A	293	ASN	CG-ND2	-5.84	1.18	1.32
1	A	293	ASN	CG-OD1	5.58	1.36	1.24

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	152	ARG	NE-CZ-NH2	5.99	123.29	120.30
1	A	173	ARG	NE-CZ-NH2	-5.98	117.31	120.30
1	A	410	ARG	NE-CZ-NH1	5.96	123.28	120.30
1	A	308	LEU	CB-CG-CD1	5.63	120.58	111.00
1	A	401	ARG	NE-CZ-NH2	-5.30	117.65	120.30

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	2105	0	2079	133	0
2	A	20	0	11	11	0
3	A	15	0	0	1	0
All	All	2140	0	2090	133	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 32.

The worst 5 of 133 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:337:GLU:O	1:A:341:VAL:HG13	1.32	1.28
1:A:338:GLU:O	1:A:341:VAL:HG22	1.31	1.26
1:A:217:GLY:HA2	1:A:218:GLU:HG2	1.22	1.10
1:A:217:GLY:CA	1:A:218:GLU:CG	2.34	1.05
1:A:150:MET:O	1:A:154:LEU:HD13	1.58	1.03

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	256/316 (81%)	232 (91%)	17 (7%)	7 (3%)	<b>6</b> <b>21</b>

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	232	GLY
1	A	201	GLN
1	A	203	ARG
1	A	228	PRO
1	A	340	TYR

### 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	224/279 (80%)	211 (94%)	13 (6%)	25 57

5 of 13 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	240	LEU
1	A	243	MET
1	A	331	LYS
1	A	226	LYS
1	A	309	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	214	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	$\# Z  > 2$
2	PNU	A	1	-	16,22,22	0.85	0	15,31,31	2.82	6 (40%)

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
2	PNU	A	1	-	-	0/7/8/8	0/2/3/3

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	1	PNU	C4-C5-N6	-5.10	119.49	125.62
2	A	1	PNU	N6-C1-N2	-3.09	120.86	126.75
2	A	1	PNU	C11-C10-N9	-2.73	120.54	122.98
2	A	1	PNU	C5-N6-C1	4.18	120.76	114.62
2	A	1	PNU	C5-C4-C3	4.52	118.62	115.39

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
2	A	1	PNU	11	0

## 5.7 Other polymers ⓘ

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	264/316 (83%)	0.17	13 (4%)	33 22	30, 45, 70, 76	1 (0%)

The worst 5 of 13 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	232	GLY	3.9
1	A	316	GLN	3.3
1	A	142	GLY	3.0
1	A	236	ILE	3.0
1	A	315	PHE	3.0

### 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(Å <sup>2</sup> )	Q<0.9
2	PNU	A	1	20/20	0.55	0.51	6.26	62,64,74,75	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.