



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 6, 2022 – 01:01 PM EDT

PDB ID : 5GXB  
Title : crystal structure of a LacY/Nanobody complex  
Authors : Jiang, X.; Wu, J.P.; Yan, N.; Kaback, H.R.  
Deposited on : 2016-09-16  
Resolution : 3.30 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

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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

MolProbity : 4.02b-467  
Xtriage (Phenix) : 1.13  
EDS : 2.31.2  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0158  
CCP4 : 7.0.044 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.31.2

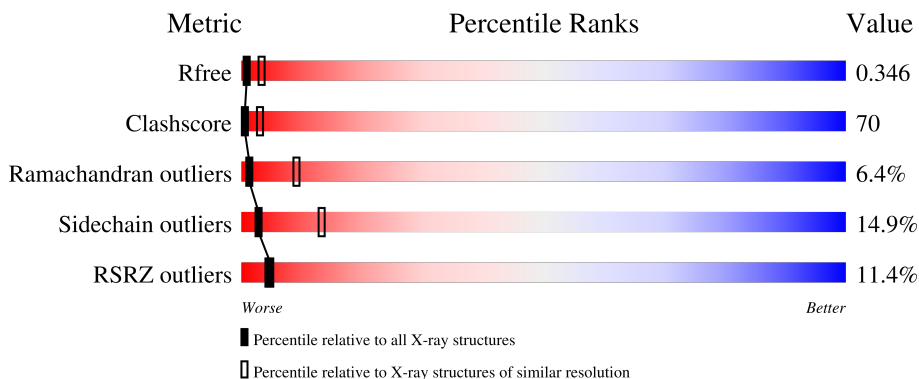
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.30 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))
$R_{free}$	130704	1149 (3.34-3.26)
Clashscore	141614	1205 (3.34-3.26)
Ramachandran outliers	138981	1183 (3.34-3.26)
Sidechain outliers	138945	1182 (3.34-3.26)
RSRZ outliers	127900	1115 (3.34-3.26)

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 of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria respectively. 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	423	
2	B	140	

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 4121 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 Lactose permease.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	387	3086	2102	465	498	21	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	46	TRP	GLY	conflict	UNP P02920
A	262	TRP	GLY	conflict	UNP P02920
A	418	HIS	-	expression tag	UNP P02920
A	419	HIS	-	expression tag	UNP P02920
A	420	HIS	-	expression tag	UNP P02920
A	421	HIS	-	expression tag	UNP P02920
A	422	HIS	-	expression tag	UNP P02920
A	423	HIS	-	expression tag	UNP P02920

- Molecule 2 is a protein called nanobody.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	132	1035	649	183	198	5	0	0	0



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 4 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	96.54Å 96.54Å 145.03Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.34 – 3.30 48.34 – 3.30	Depositor EDS
% Data completeness (in resolution range)	95.2 (48.34-3.30) 95.2 (48.34-3.30)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.20 (at 3.33Å)	Xtrriage
Refinement program	PHENIX (dev_2405: ???)	Depositor
R, $R_{free}$	0.340 , 0.346 0.339 , 0.346	Depositor DCC
$R_{free}$ test set	513 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	93.7	Xtrriage
Anisotropy	0.076	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 37.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.43$ , $\langle L^2 \rangle = 0.26$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.79	EDS
Total number of atoms	4121	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	74.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.29% 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 [i](#)

### 5.1 Standard geometry [i](#)

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	0.57	2/3182 (0.1%)	0.74	15/4312 (0.3%)
2	B	0.72	1/1059 (0.1%)	0.80	0/1433
All	All	0.61	3/4241 (0.1%)	0.76	15/5745 (0.3%)

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

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	398	PHE	C-N	8.98	1.54	1.34
2	B	102	PRO	N-CD	5.21	1.55	1.47
1	A	31	PRO	N-CD	5.10	1.54	1.47

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	92	ILE	CB-CA-C	12.58	136.76	111.60
1	A	156	SER	O-C-N	-11.91	103.64	122.70
1	A	87	PHE	CB-CA-C	10.88	132.16	110.40
1	A	170	PHE	CB-CA-C	-9.02	92.36	110.40
1	A	93	PHE	CB-CA-C	8.35	127.11	110.40

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	156	SER	Mainchain
1	A	87	PHE	Peptide

## 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	3086	0	3120	377	1
2	B	1035	0	988	216	1
All	All	4121	0	4108	572	1

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 70.

The worst 5 of 572 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:286:ILE:CD1	1:A:290:ASN:HB2	1.26	1.64
2:B:0:MET:CE	2:B:101:ARG:HD3	1.26	1.59
1:A:286:ILE:HD11	1:A:290:ASN:CB	1.36	1.55
1:A:247:PHE:HE1	1:A:251:PHE:CE1	1.37	1.41
2:B:22:CYS:SG	2:B:96:CYS:SG	1.42	1.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:39:HIS:O	2:B:130:HIS:NE2[7_556]	1.83	0.37

## 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	383/423 (90%)	339 (88%)	34 (9%)	10 (3%)	5	27
2	B	130/140 (93%)	89 (68%)	18 (14%)	23 (18%)	0	1
All	All	513/563 (91%)	428 (83%)	52 (10%)	33 (6%)	1	9

5 of 33 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	39	HIS
1	A	240	ASP
1	A	252	ALA
1	A	313	LEU
2	B	27	ARG

### 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	325/354 (92%)	292 (90%)	33 (10%)	7	27
2	B	105/112 (94%)	74 (70%)	31 (30%)	0	1
All	All	430/466 (92%)	366 (85%)	64 (15%)	3	13

5 of 64 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	103	SER
2	B	111	THR
1	A	286	ILE
1	A	259	ARG
2	B	112	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:



Mol	Chain	Res	Type
2	B	84	ASN
2	B	124	GLN
2	B	131	HIS
1	A	242	GLN
2	B	3	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 monosaccharides in this entry.

### 5.6 Ligand geometry [i](#)

There are no ligands in this entry.

### 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	387/423 (91%)	0.58	32 (8%) 11 11	55, 69, 91, 108	1 (0%)
2	B	132/140 (94%)	1.29	27 (20%) 1 1	69, 81, 99, 107	0
All	All	519/563 (92%)	0.76	59 (11%) 5 4	55, 73, 94, 108	1 (0%)

The worst 5 of 59 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	101	TYR	8.7
1	A	342	GLU	7.9
2	B	105	SER	6.9
2	B	39	GLN	6.9
2	B	106	GLY	5.4

### 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 monosaccharides in this entry.

### 6.4 Ligands [i](#)

There are no ligands in this entry.

### 6.5 Other polymers [i](#)

There are no such residues in this entry.