



# wwPDB X-ray Structure Validation Summary Report

May 21, 2020 – 04:17 pm BST

PDB ID : 5N8B  
Title : CRYSTAL STRUCTURE OF STREPTAVIDIN WITH PEPTIDE AFPDY-LAEYHGG  
Authors : Lyamichev, V.; Goodrich, L.; Sullivan, E.; Bannen, R.; Benz, J.; Albert, T.; Patel, J.  
Deposited on : 2017-02-23  
Resolution : 1.03 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 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 : 2.11  
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) : 2.11

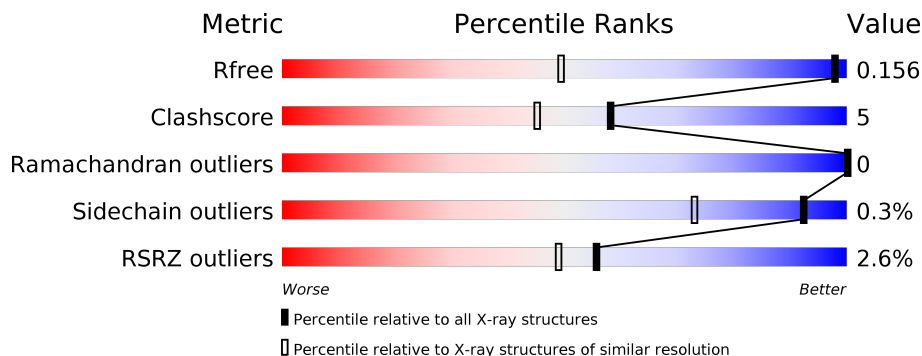
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.03 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	1596 (1.10-0.98)
Clashscore	141614	1677 (1.10-0.98)
Ramachandran outliers	138981	1591 (1.10-0.98)
Sidechain outliers	138945	1589 (1.10-0.98)
RSRZ outliers	127900	1557 (1.10-0.98)

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 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	183	
1	B	183	
1	D	183	
1	G	183	
2	C	13	
2	E	13	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	F	13	 100%
2	H	13	 92% 8%

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 5153 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 Streptavidin.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
1	A	121	Total 948	C 601	N 159	O 188	0	10	0
1	B	120	Total 941	C 596	N 157	O 188	0	10	0
1	D	125	Total 973	C 612	N 166	O 195	0	10	0
1	G	121	Total 937	C 592	N 158	O 187	0	8	0

- Molecule 2 is a protein called ALA-PHE-PRO-ASP-TYR-LEU-ALA-GLU-TYR-HIS-GLY-GLY-NH2.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
2	E	13	Total 96	C 63	N 15	O 18	0	0	1
2	C	13	Total 96	C 63	N 15	O 18	0	0	1
2	F	13	Total 96	C 63	N 15	O 18	0	0	1
2	H	13	Total 96	C 63	N 15	O 18	0	0	1

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	230	Total 230	O 230	0	0
3	E	33	Total 33	O 33	0	0
3	B	179	Total 182	O 182	0	3
3	C	26	Total 26	O 26	0	0

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
3	D	209	Total 211	O 211	0	2
3	F	33	Total 33	O 33	0	0
3	G	230	Total 231	O 231	0	1
3	H	24	Total 24	O 24	0	0



GLY  
VAL  
ASN  
ASN  
GLY  
ASN  
PRO  
LEU  
ASP  
ALA  
VAL  
GLN  
GLN

- Molecule 2: ALA-PHE-PRO-ASP-TYR-LEU-ALA-GLU-TYR-HIS-GLY-GLY-NH2

Chain E:  92% 8%

11  
612  
713

- Molecule 2: ALA-PHE-PRO-ASP-TYR-LEU-ALA-GLU-TYR-HIS-GLY-GLY-NH2

Chain C:  100%

There are no outlier residues recorded for this chain.

- Molecule 2: ALA-PHE-PRO-ASP-TYR-LEU-ALA-GLU-TYR-HIS-GLY-GLY-NH2

Chain F:  100%

There are no outlier residues recorded for this chain.

- Molecule 2: ALA-PHE-PRO-ASP-TYR-LEU-ALA-GLU-TYR-HIS-GLY-GLY-NH2

Chain H:  92% 8%

11  
612  
713

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	51.32Å 65.76Å 78.13Å 90.00° 103.72° 90.00°	Depositor
Resolution (Å)	75.90 – 1.03 49.85 – 1.03	Depositor EDS
% Data completeness (in resolution range)	99.7 (75.90-1.03) 99.7 (49.85-1.03)	Depositor EDS
$R_{merge}$	0.04	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.46 (at 1.03Å)	Xtrriage
Refinement program	REFMAC 5.8.0158	Depositor
R, $R_{free}$	0.138 , 0.155 0.139 , 0.156	Depositor DCC
$R_{free}$ test set	12218 reflections (4.93%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	10.7	Xtrriage
Anisotropy	0.293	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 35.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.98	EDS
Total number of atoms	5153	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	16.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.80% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: NH2

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	0.41	0/999	0.71	0/1368
1	B	0.39	0/992	0.71	0/1360
1	D	0.41	0/1024	0.68	0/1402
1	G	0.38	0/982	0.69	0/1346
2	C	0.43	0/99	0.67	0/134
2	E	0.37	0/99	0.72	0/134
2	F	0.41	0/99	0.68	0/134
2	H	0.40	0/99	0.70	0/134
All	All	0.40	0/4393	0.70	0/6012

There are no bond length outliers.

There are no bond angle outliers.

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	948	0	921	15	0
1	B	941	0	910	14	0
1	D	973	0	938	12	0
1	G	937	0	899	8	0
2	C	96	0	80	0	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	E	96	0	80	1	0
2	F	96	0	80	0	0
2	H	96	0	80	1	0
3	A	230	0	0	3	0
3	B	182	0	0	6	0
3	C	26	0	0	0	0
3	D	211	0	0	10	1
3	E	33	0	0	0	0
3	F	33	0	0	0	0
3	G	231	0	0	0	0
3	H	24	0	0	0	0
All	All	5153	0	3988	39	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 5.

The worst 5 of 39 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:D:107[A]:GLN:OE1	3:D:201[A]:HOH:O	1.60	1.19
1:B:107[B]:GLN:OE1	3:B:201[B]:HOH:O	1.62	1.17
1:A:103:ARG:HH21	1:A:129[B]:THR:CG2	1.71	1.01
1:A:103:ARG:HH21	1:A:129[B]:THR:HG21	1.23	0.98
1:D:107[A]:GLN:CG	3:D:201[A]:HOH:O	2.16	0.94

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)
3:D:202:HOH:O	3:D:203:HOH:O 2_546]	2.12	0.08

## 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	128/183 (70%)	128 (100%)	0	0	100	100
1	B	127/183 (69%)	125 (98%)	2 (2%)	0	100	100
1	D	132/183 (72%)	128 (97%)	4 (3%)	0	100	100
1	G	126/183 (69%)	124 (98%)	2 (2%)	0	100	100
2	C	11/13 (85%)	11 (100%)	0	0	100	100
2	E	11/13 (85%)	11 (100%)	0	0	100	100
2	F	11/13 (85%)	11 (100%)	0	0	100	100
2	H	11/13 (85%)	11 (100%)	0	0	100	100
All	All	557/784 (71%)	549 (99%)	8 (1%)	0	100	100

There are no Ramachandran outliers to report.

### 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	98/134 (73%)	97 (99%)	1 (1%)	76	45
1	B	98/134 (73%)	98 (100%)	0	100	100
1	D	100/134 (75%)	100 (100%)	0	100	100
1	G	96/134 (72%)	96 (100%)	0	100	100
2	C	8/8 (100%)	8 (100%)	0	100	100
2	E	8/8 (100%)	8 (100%)	0	100	100
2	F	8/8 (100%)	8 (100%)	0	100	100
2	H	8/8 (100%)	8 (100%)	0	100	100
All	All	424/568 (75%)	423 (100%)	1 (0%)	92	74

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	116	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	82	ASN
1	B	82	ASN
1	D	82	ASN
1	G	82	ASN

### 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](#)

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	121/183 (66%)	-0.27	1 (0%) 86 81	7, 10, 19, 28	2 (1%)
1	B	120/183 (65%)	-0.20	5 (4%) 36 30	7, 12, 25, 34	2 (1%)
1	D	125/183 (68%)	-0.08	5 (4%) 38 32	7, 12, 24, 32	2 (1%)
1	G	121/183 (66%)	-0.29	3 (2%) 57 50	7, 12, 26, 36	2 (1%)
2	C	12/13 (92%)	-0.46	0 100 100	8, 11, 13, 15	0
2	E	12/13 (92%)	-0.42	0 100 100	8, 10, 13, 15	0
2	F	12/13 (92%)	-0.29	0 100 100	10, 15, 16, 18	0
2	H	12/13 (92%)	-0.47	0 100 100	8, 12, 14, 18	0
All	All	535/784 (68%)	-0.23	14 (2%) 56 49	7, 12, 23, 36	8 (1%)

The worst 5 of 14 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	15	ALA	6.9
1	G	15	ALA	6.0
1	D	137	ALA	5.6
1	D	138	ALA	5.1
1	D	15	ALA	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

There are no ligands in this entry.

## 6.5 Other polymers

There are no such residues in this entry.