



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 09:10 AM GMT

PDB ID : 3HG0  
Title : Crystal structure of a DARPin in complex with ORF49 from Lactococcal phage TP901-1  
Authors : Veesler, D.; Dreier, B.; Blangy, S.; Lichiere, J.; Tremblay, D.; Moineau, S.; Spinelli, S.; Tegoni, M.; Pluckthun, A.; Campanacci, V.; Cambillau, C.  
Deposited on : 2009-05-13  
Resolution : 2.10 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

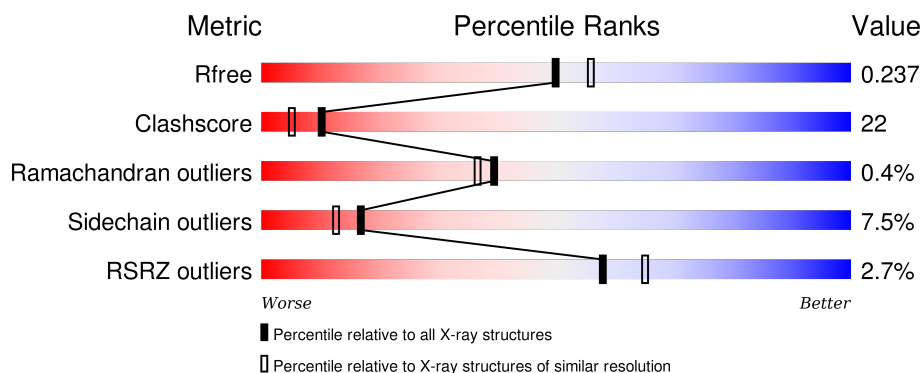
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.10 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	3939 (2.10-2.10)
Clashscore	102246	4460 (2.10-2.10)
Ramachandran outliers	100387	4413 (2.10-2.10)
Sidechain outliers	100360	4414 (2.10-2.10)
RSRZ outliers	91569	3948 (2.10-2.10)

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	169	<div> <div>3%</div> <div>64%</div> <div>14%</div> <div>• •</div> <div>19%</div> </div>
1	B	169	<div> <div>52%</div> <div>21%</div> <div>•</div> <div>23%</div> </div>
1	C	169	<div> <div>3%</div> <div>57%</div> <div>18%</div> <div>•</div> <div>22%</div> </div>
2	D	136	<div> <div>3%</div> <div>59%</div> <div>26%</div> <div>7%</div> <div>9%</div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 4379 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 Baseplate protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	137	Total	C	N	O	S	0	3	0
			1035	646	189	196	4			
1	B	130	Total	C	N	O	S	0	3	0
			977	610	172	191	4			
1	C	132	Total	C	N	O	S	0	3	0
			989	618	177	190	4			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	164	HIS	-	EXPRESSION TAG	UNP Q9G096
A	165	HIS	-	EXPRESSION TAG	UNP Q9G096
A	166	HIS	-	EXPRESSION TAG	UNP Q9G096
A	167	HIS	-	EXPRESSION TAG	UNP Q9G096
A	168	HIS	-	EXPRESSION TAG	UNP Q9G096
A	169	HIS	-	EXPRESSION TAG	UNP Q9G096
B	164	HIS	-	EXPRESSION TAG	UNP Q9G096
B	165	HIS	-	EXPRESSION TAG	UNP Q9G096
B	166	HIS	-	EXPRESSION TAG	UNP Q9G096
B	167	HIS	-	EXPRESSION TAG	UNP Q9G096
B	168	HIS	-	EXPRESSION TAG	UNP Q9G096
B	169	HIS	-	EXPRESSION TAG	UNP Q9G096
C	164	HIS	-	EXPRESSION TAG	UNP Q9G096
C	165	HIS	-	EXPRESSION TAG	UNP Q9G096
C	166	HIS	-	EXPRESSION TAG	UNP Q9G096
C	167	HIS	-	EXPRESSION TAG	UNP Q9G096
C	168	HIS	-	EXPRESSION TAG	UNP Q9G096
C	169	HIS	-	EXPRESSION TAG	UNP Q9G096

- Molecule 2 is a protein called Designed Ankyrin Repeat Protein (DARPin) 20.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	124	Total	C	N	O	S	0	0	0
			932	583	159	187	3			

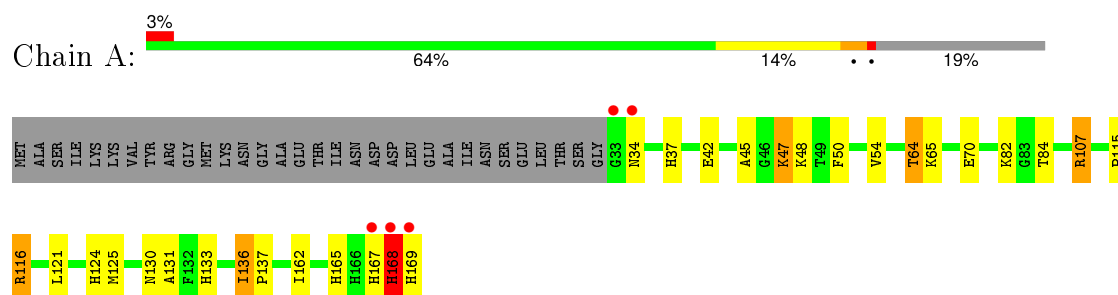
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	158	Total	O	0	0
			158	158		
3	B	93	Total	O	0	0
			93	93		
3	C	107	Total	O	0	0
			107	107		
3	D	88	Total	O	0	0
			88	88		

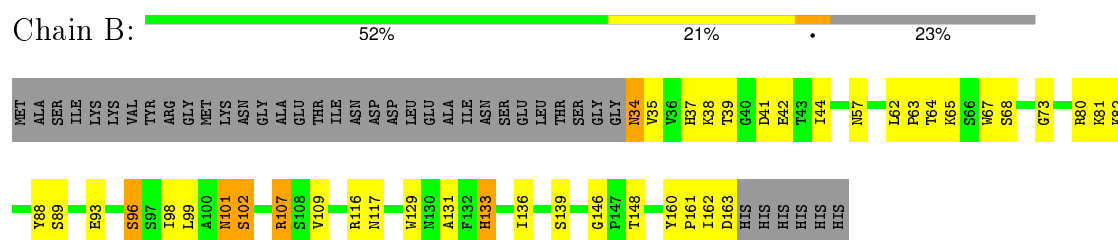
### 3 Residue-property plots [i](#)

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.

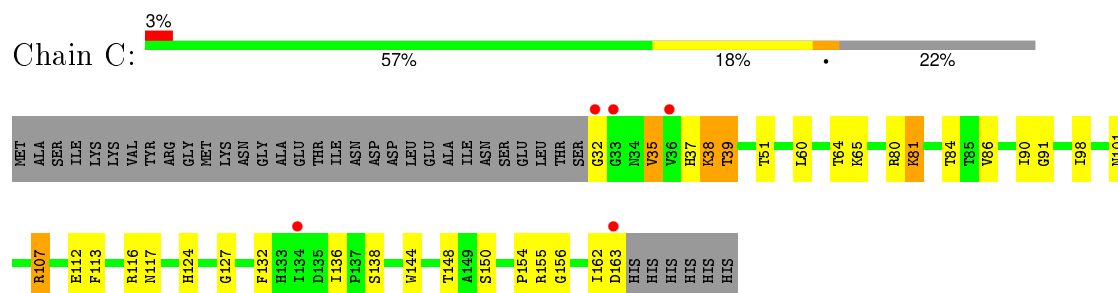
- Molecule 1: Baseplate protein



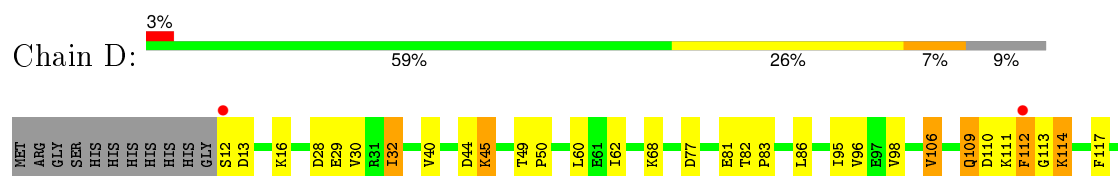
- Molecule 1: Baseplate protein



- Molecule 1: Baseplate protein



- Molecule 2: Designed Ankyrin Repeat Protein (DARPin) 20



S120	I121	D122	N123	G124	N125	E126	D127	L128	A129	E130	I131	L132	Q133	K134	L135	ASN

•

•

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	46.65Å 80.44Å 182.87Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.89 – 2.10 19.89 – 2.10	Depositor EDS
% Data completeness (in resolution range)	97.0 (19.89-2.10) 97.0 (19.89-2.10)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.55 (at 2.09Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE: 1.4_56)	Depositor
R, $R_{free}$	0.208 , 0.243 0.202 , 0.237	Depositor DCC
$R_{free}$ test set	996 reflections (2.50%)	DCC
Wilson B-factor (Å <sup>2</sup> )	33.3	Xtriage
Anisotropy	0.483	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 48.2	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 39881 reflections	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	4379	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	50.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.07% 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

### 5.1 Standard geometry

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	0.62	0/1071	0.82	2/1456 (0.1%)
1	B	0.61	0/1001	0.79	2/1361 (0.1%)
1	C	0.63	0/1016	0.85	2/1377 (0.1%)
2	D	0.56	0/941	0.73	0/1273
All	All	0.61	0/4029	0.80	6/5467 (0.1%)

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(°)
1	B	133	HIS	N-CA-CB	7.91	124.84	110.60
1	A	133	HIS	N-CA-CB	6.47	122.25	110.60
1	B	146	GLY	N-CA-C	-5.99	98.13	113.10
1	C	81[A]	LYS	CB-CA-C	-5.78	98.83	110.40
1	C	81[B]	LYS	CB-CA-C	-5.78	98.83	110.40

There are no chirality outliers.

There are no planarity outliers.

### 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	1035	0	1011	31	0
1	B	977	0	965	58	0
1	C	989	0	985	59	0
2	D	932	0	933	41	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	158	0	0	2	0
3	B	93	0	0	8	0
3	C	107	0	0	7	0
3	D	88	0	0	4	0
All	All	4379	0	3894	170	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 22.

The worst 5 of 170 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:C:81[A]:LYS:NZ	1:C:112:GLU:HG3	1.08	1.36
1:C:81[A]:LYS:NZ	1:C:112:GLU:CG	1.97	1.27
1:C:81[A]:LYS:HZ3	1:C:112:GLU:CG	1.55	1.12
2:D:45:LYS:HD2	2:D:45:LYS:H	1.11	1.10
1:B:107:ARG:HG2	1:B:107:ARG:HH11	1.05	1.08

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	138/169 (82%)	134 (97%)	2 (1%)	2 (1%)	14	7
1	B	131/169 (78%)	125 (95%)	6 (5%)	0	100	100
1	C	133/169 (79%)	126 (95%)	7 (5%)	0	100	100
2	D	122/136 (90%)	115 (94%)	7 (6%)	0	100	100
All	All	524/643 (82%)	500 (95%)	22 (4%)	2 (0%)	39	37

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	168	HIS
1	A	137	PRO

### 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	116/139 (84%)	109 (94%)	7 (6%)	24	20
1	B	110/139 (79%)	103 (94%)	7 (6%)	22	18
1	C	110/139 (79%)	102 (93%)	8 (7%)	17	13
2	D	98/108 (91%)	85 (87%)	13 (13%)	5	2
All	All	434/525 (83%)	399 (92%)	35 (8%)	17	10

5 of 35 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	38	LYS
1	C	136	ILE
2	D	123	ASN
1	C	39	THR
1	C	107[A]	ARG

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

Mol	Chain	Res	Type
1	B	57	ASN
1	B	117	ASN
1	C	57	ASN
1	B	34	ASN
1	C	34	ASN

### 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](#)

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	137/169 (81%)	-0.32	5 (3%) 46 55	27, 37, 70, 114	0
1	B	130/169 (76%)	-0.13	0 100 100	32, 51, 75, 105	0
1	C	132/169 (78%)	-0.20	5 (3%) 44 53	32, 47, 75, 103	0
2	D	124/136 (91%)	-0.02	4 (3%) 51 60	23, 47, 81, 137	1 (0%)
All	All	523/643 (81%)	-0.17	14 (2%) 58 65	23, 46, 77, 137	1 (0%)

The worst 5 of 14 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	32	GLY	6.1
2	D	135	LEU	5.9
1	A	169	HIS	4.3
2	D	134	LYS	4.2
2	D	12	SER	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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.