



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 3, 2022 – 04:51 PM EST

PDB ID : 7N2B  
Title : A DARPin semi-rigidly fused to the 3TEL crystallization chaperone  
Authors : Sarath Nawarathange, S.D.; Gajjar, P.; Bunn, D.; Stewart, C.; Doukov, T.;  
Moody, J.D.  
Deposited on : 2021-05-28  
Resolution : 3.22 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	:	2.27
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)	:	2.27

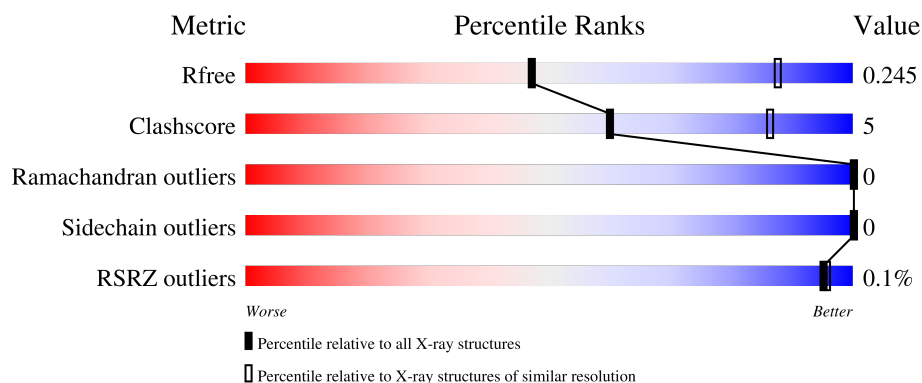
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*



The reported resolution of this entry is 3.22 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	1335 (3.24-3.20)
Clashscore	141614	1460 (3.24-3.20)
Ramachandran outliers	138981	1437 (3.24-3.20)
Sidechain outliers	138945	1436 (3.24-3.20)
RSRZ outliers	127900	1291 (3.24-3.20)

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	416	 85% 12% •
1	B	416	 87% 11% •

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 6067 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 Transcription factor ETV6,Transcription factor ETV6,Transcription factor ETV6,3TEL-rigid-DARPin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	402	Total	C	N	O	S	0	0	0
			3033	1930	533	566	4			
1	B	404	Total	C	N	O	S	0	0	0
			3034	1928	522	580	4			

There are 38 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	GLY	-	expression tag	UNP P41212
A	66	GLU	VAL	engineered mutation	UNP P41212
A	76	LYS	-	linker	UNP P41212
A	77	GLN	-	linker	UNP P41212
A	78	ARG	-	linker	UNP P41212
A	79	PRO	-	linker	UNP P41212
A	80	GLY	-	linker	UNP P41212
A	81	GLY	-	linker	UNP P41212
A	82	GLY	-	linker	UNP P41212
A	83	GLY	-	linker	UNP P41212
A	84	SER	-	linker	UNP P41212
A	85	THR	-	linker	UNP P41212
A	164	PRO	-	linker	UNP P41212
A	165	GLY	-	linker	UNP P41212
A	166	GLY	-	linker	UNP P41212
A	167	GLY	-	linker	UNP P41212
A	168	GLY	-	linker	UNP P41212
A	169	SER	-	linker	UNP P41212
A	170	THR	-	linker	UNP P41212
B	0	GLY	-	expression tag	UNP P41212
B	66	GLU	VAL	engineered mutation	UNP P41212
B	76	LYS	-	linker	UNP P41212
B	77	GLN	-	linker	UNP P41212
B	78	ARG	-	linker	UNP P41212

*Continued on next page...*

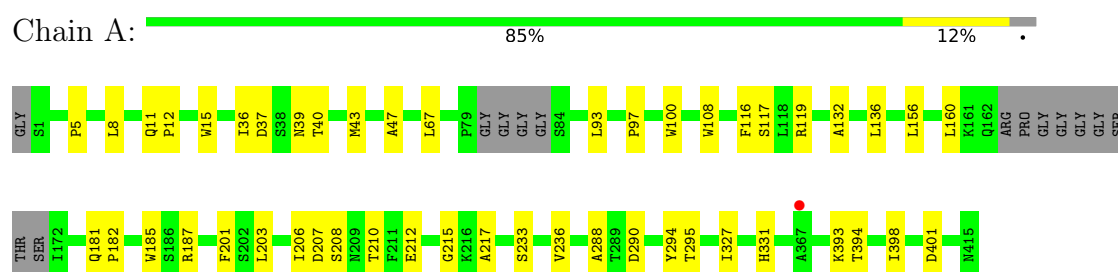
*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	79	PRO	-	linker	UNP P41212
B	80	GLY	-	linker	UNP P41212
B	81	GLY	-	linker	UNP P41212
B	82	GLY	-	linker	UNP P41212
B	83	GLY	-	linker	UNP P41212
B	84	SER	-	linker	UNP P41212
B	85	THR	-	linker	UNP P41212
B	164	PRO	-	linker	UNP P41212
B	165	GLY	-	linker	UNP P41212
B	166	GLY	-	linker	UNP P41212
B	167	GLY	-	linker	UNP P41212
B	168	GLY	-	linker	UNP P41212
B	169	SER	-	linker	UNP P41212
B	170	THR	-	linker	UNP P41212

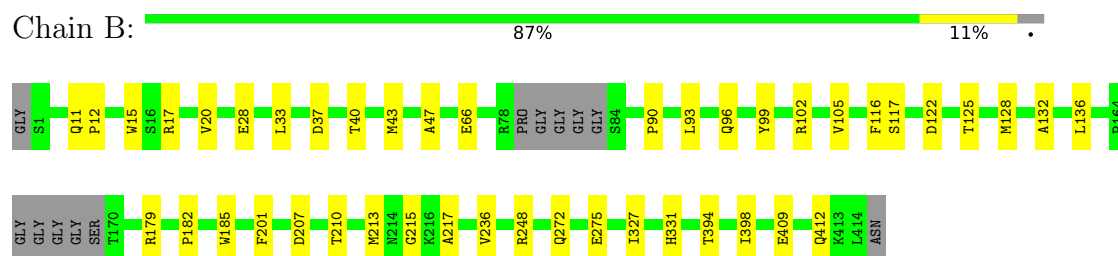
### 3 Residue-property plots

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes 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: Transcription factor ETV6,Transcription factor ETV6,Transcription factor ETV6,3 TEL-rigid-DARPin



- Molecule 1: Transcription factor ETV6,Transcription factor ETV6,Transcription factor ETV6,3 TEL-rigid-DARPin



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	45.96Å 63.62Å 166.00Å 90.00° 90.16° 90.00°	Depositor
Resolution (Å)	40.16 – 3.22 83.00 – 3.22	Depositor EDS
% Data completeness (in resolution range)	87.3 (40.16-3.22) 78.2 (83.00-3.22)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.06 (at 3.19Å)	Xtriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, $R_{free}$	0.226 , 0.242 0.225 , 0.245	Depositor DCC
$R_{free}$ test set	681 reflections (4.93%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	76.9	Xtriage
Anisotropy	0.897	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 32.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.438 for h,-k,-l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	6067	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	92.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.15% 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

### 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	0.26	0/3103	0.44	0/4238
1	B	0.26	0/3103	0.45	0/4245
All	All	0.26	0/6206	0.45	0/8483

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

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	3033	0	2810	27	0
1	B	3034	0	2759	27	0
All	All	6067	0	5569	53	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 5.

The worst 5 of 53 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:B:47:ALA:HB2	1:B:116:PHE:HB3	1.72	0.70
1:B:43:MET:HA	1:B:117:SER:O	1.94	0.67

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:43:MET:HA	1:A:117:SER:O	1.95	0.66
1:A:185:TRP:HB2	1:A:215:GLY:HA3	1.80	0.62
1:A:187:ARG:HH21	1:A:212:GLU:HG2	1.63	0.62

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	396/416 (95%)	390 (98%)	6 (2%)	0	100	100
1	B	398/416 (96%)	391 (98%)	7 (2%)	0	100	100
All	All	794/832 (95%)	781 (98%)	13 (2%)	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	289/348 (83%)	289 (100%)	0	100	100
1	B	286/348 (82%)	286 (100%)	0	100	100
All	All	575/696 (83%)	575 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.



Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

## 5.6 Ligand geometry ⓘ

There are no ligands in this entry.

## 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	402/416 (96%)	-0.22	1 (0%) 95 94	66, 93, 120, 148	0
1	B	404/416 (97%)	-0.23	0 100 100	65, 93, 119, 145	0
All	All	806/832 (96%)	-0.23	1 (0%) 95 96	65, 93, 120, 148	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	367	ALA	2.2

### 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.