



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 10:02 AM GMT

PDB ID : 3KNQ  
Title : Beta Turn Optimization of the Gene-3-Protein of Filamentous Phage Fd  
Authors : Jakob, R.P; Dobbek, H.  
Deposited on : 2009-11-12  
Resolution : 2.13 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

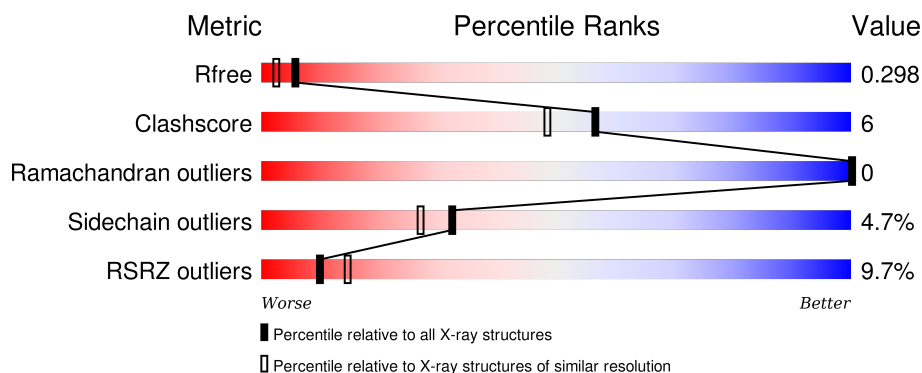
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.13 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	1693 (2.16-2.12)
Clashscore	102246	1824 (2.16-2.12)
Ramachandran outliers	100387	1798 (2.16-2.12)
Sidechain outliers	100360	1798 (2.16-2.12)
RSRZ outliers	91569	1699 (2.16-2.12)

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	223	 10% 70% 13% • 15%
1	B	223	 6% 68% 13% 18%

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 3196 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 Attachment protein G3P.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	189	Total	C	N	O	S	0	0	0
			1498	968	238	290	2			
1	B	183	Total	C	N	O	S	0	0	0
			1454	940	231	281	2			

There are 82 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	INITIATING METHIONINE	UNP P03661
A	?	-	CYS	DELETION	UNP P03661
A	11	SER	PRO	ENGINEERED	UNP P03661
A	13	ILE	THR	ENGINEERED	UNP P03661
A	15	GLY	ASN	ENGINEERED	UNP P03661
A	29	TRP	ARG	ENGINEERED	UNP P03661
A	36	ILE	CYS	ENGINEERED	UNP P03661
A	39	LYS	ASN	ENGINEERED	UNP P03661
A	46	ILE	CYS	ENGINEERED	UNP P03661
A	53	VAL	CYS	ENGINEERED	UNP P03661
A	55	ALA	GLY	ENGINEERED	UNP P03661
A	56	ILE	THR	ENGINEERED	UNP P03661
A	60	VAL	ILE	ENGINEERED	UNP P03661
A	101	ILE	THR	ENGINEERED	UNP P03661
A	129	HIS	GLN	ENGINEERED	UNP P03661
A	138	GLY	ASN	ENGINEERED	UNP P03661
A	144	VAL	ARG	ENGINEERED	UNP P03661
A	145	ASN	GLN	ENGINEERED	UNP P03661
A	147	VAL	ALA	ENGINEERED	UNP P03661
A	?	-	GLN	DELETION	UNP P03661
A	?	-	GLY	DELETION	UNP P03661
A	?	-	THR	DELETION	UNP P03661
A	?	-	ASP	DELETION	UNP P03661
A	?	-	PRO	DELETION	UNP P03661
A	?	-	VAL	DELETION	UNP P03661

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	157	VAL	-	INSERTION	UNP P03661
A	158	ASN	-	INSERTION	UNP P03661
A	159	GLY	-	INSERTION	UNP P03661
A	185	VAL	CYS	ENGINEERED	UNP P03661
A	196	LEU	PHE	ENGINEERED	UNP P03661
A	198	ALA	CYS	ENGINEERED	UNP P03661
A	204	LEU	SER	ENGINEERED	UNP P03661
A	206	TYR	ASP	ENGINEERED	UNP P03661
A	215	PRO	GLY	ENGINEERED	UNP P03661
A	216	SER	GLY	ENGINEERED	UNP P03661
A	218	HIS	-	EXPRESSION TAG	UNP P03661
A	219	HIS	-	EXPRESSION TAG	UNP P03661
A	220	HIS	-	EXPRESSION TAG	UNP P03661
A	221	HIS	-	EXPRESSION TAG	UNP P03661
A	222	HIS	-	EXPRESSION TAG	UNP P03661
A	223	HIS	-	EXPRESSION TAG	UNP P03661
B	1	MET	-	INITIATING METHIONINE	UNP P03661
B	?	-	CYS	DELETION	UNP P03661
B	11	SER	PRO	ENGINEERED	UNP P03661
B	13	ILE	THR	ENGINEERED	UNP P03661
B	15	GLY	ASN	ENGINEERED	UNP P03661
B	29	TRP	ARG	ENGINEERED	UNP P03661
B	36	ILE	CYS	ENGINEERED	UNP P03661
B	39	LYS	ASN	ENGINEERED	UNP P03661
B	46	ILE	CYS	ENGINEERED	UNP P03661
B	53	VAL	CYS	ENGINEERED	UNP P03661
B	55	ALA	GLY	ENGINEERED	UNP P03661
B	56	ILE	THR	ENGINEERED	UNP P03661
B	60	VAL	ILE	ENGINEERED	UNP P03661
B	101	ILE	THR	ENGINEERED	UNP P03661
B	129	HIS	GLN	ENGINEERED	UNP P03661
B	138	GLY	ASN	ENGINEERED	UNP P03661
B	144	VAL	ARG	ENGINEERED	UNP P03661
B	145	ASN	GLN	ENGINEERED	UNP P03661
B	147	VAL	ALA	ENGINEERED	UNP P03661
B	?	-	GLN	DELETION	UNP P03661
B	?	-	GLY	DELETION	UNP P03661
B	?	-	THR	DELETION	UNP P03661
B	?	-	ASP	DELETION	UNP P03661
B	?	-	PRO	DELETION	UNP P03661
B	?	-	VAL	DELETION	UNP P03661
B	157	VAL	-	INSERTION	UNP P03661

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	158	ASN	-	INSERTION	UNP P03661
B	159	GLY	-	INSERTION	UNP P03661
B	185	VAL	CYS	ENGINEERED	UNP P03661
B	196	LEU	PHE	ENGINEERED	UNP P03661
B	198	ALA	CYS	ENGINEERED	UNP P03661
B	204	LEU	SER	ENGINEERED	UNP P03661
B	206	TYR	ASP	ENGINEERED	UNP P03661
B	215	PRO	GLY	ENGINEERED	UNP P03661
B	216	SER	GLY	ENGINEERED	UNP P03661
B	218	HIS	-	EXPRESSION TAG	UNP P03661
B	219	HIS	-	EXPRESSION TAG	UNP P03661
B	220	HIS	-	EXPRESSION TAG	UNP P03661
B	221	HIS	-	EXPRESSION TAG	UNP P03661
B	222	HIS	-	EXPRESSION TAG	UNP P03661
B	223	HIS	-	EXPRESSION TAG	UNP P03661

- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	129	Total O 129 129	0	0
2	B	115	Total O 115 115	0	0



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	56.51Å 86.97Å 96.29Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	33.77 – 2.13 33.77 – 2.13	Depositor EDS
% Data completeness (in resolution range)	96.4 (33.77-2.13) 96.5 (33.77-2.13)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.46 (at 2.14Å)	Xtriage
Refinement program	BUSTER 2.8.0	Depositor
R, $R_{free}$	0.214 , 0.271 0.239 , 0.298	Depositor DCC
$R_{free}$ test set	1324 reflections (5.04%)	DCC
Wilson B-factor (Å <sup>2</sup> )	36.8	Xtriage
Anisotropy	0.329	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 61.3	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.40$ , $\langle L^2 \rangle = 0.22$	Xtriage
Outliers	0 of 26294 reflections	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	3196	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	53.0	wwPDB-VP

Xtriage'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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.73	0/1549	0.73	2/2128 (0.1%)
1	B	0.67	0/1505	0.70	0/2069
All	All	0.70	0/3054	0.72	2/4197 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	160	LYS	N-CA-CB	-7.85	96.47	110.60
1	A	159	GLY	N-CA-C	5.03	125.66	113.10

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	1498	0	1405	18	0
1	B	1454	0	1356	21	0
2	A	129	0	0	0	0
2	B	115	0	0	0	0
All	All	3196	0	2761	35	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 6.

The worst 5 of 35 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:18:THR:H	1:B:213:ASN:HD21	1.18	0.89
1:A:18:THR:H	1:A:213:ASN:HD21	1.19	0.88
1:B:46:ILE:CD1	1:B:102:TYR:HB2	2.04	0.88
1:A:203:GLN:HE22	1:B:43:VAL:H	1.34	0.75
1:A:43:VAL:H	1:B:203:GLN:HE22	1.35	0.74

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	185/223 (83%)	180 (97%)	5 (3%)	0	100	100
1	B	179/223 (80%)	176 (98%)	3 (2%)	0	100	100
All	All	364/446 (82%)	356 (98%)	8 (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	162/182 (89%)	154 (95%)	8 (5%)	31	25
1	B	157/182 (86%)	150 (96%)	7 (4%)	34	29

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	319/364 (88%)	304 (95%)	15 (5%)	32 27

5 of 15 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	156	THR
1	A	213	ASN
1	B	156	THR
1	A	116	GLN
1	B	116	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	213	ASN
1	B	12	HIS
1	B	188	HIS
1	A	203	GLN
1	B	164	GLN

### 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 carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

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	189/223 (84%)	0.57	22 (11%) 6 9	29, 49, 94, 114	0
1	B	183/223 (82%)	0.52	14 (7%) 16 22	24, 52, 80, 115	0
All	All	372/446 (83%)	0.55	36 (9%) 10 14	24, 51, 89, 115	0

The worst 5 of 36 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	64	ILE	6.2
1	A	157	VAL	4.8
1	A	37	LEU	4.4
1	A	62	LEU	4.1
1	A	101	ILE	3.5

### 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.