



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 14, 2018 – 12:48 PM EDT

PDB ID : 5N47
Title : Structure of Anticalin N7E in complex with the three-domain fragment Fn7B8 of human oncofetal fibronectin
Authors : Schiefner, A.; Skerra, A.
Deposited on : 2017-02-10
Resolution : 3.00 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

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
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030736
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	rb-20030736

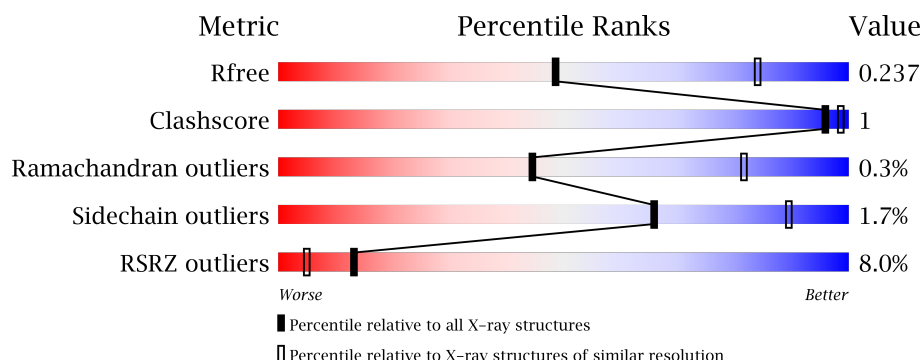
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.00 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}	100719	1692 (3.00-3.00)
Clashscore	112137	2037 (3.00-3.00)
Ramachandran outliers	110173	1973 (3.00-3.00)
Sidechain outliers	110143	1976 (3.00-3.00)
RSRZ outliers	101464	1716 (3.00-3.00)

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 ≥ 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	188	<div> <div>89%</div> <div>6%</div> <div>.</div> </div>
1	C	188	<div> <div>%</div> <div>85%</div> <div>9%</div> <div>6%</div> </div>
1	E	188	<div> <div>2%</div> <div>88%</div> <div>..</div> <div>9%</div> </div>
2	B	284	<div> <div>61%</div> <div>36%</div> </div>
2	D	284	<div> <div>6%</div> <div>94%</div> <div>..</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	284	<div> <div>28%</div> <div>94%</div> <div>..</div> </div>

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 9943 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 Neutrophil gelatinase-associated lipocalin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	180	Total	C	N	O	S	0	0	0
			1463	943	250	266	4			
1	C	176	Total	C	N	O	S	0	0	0
			1436	925	245	262	4			
1	E	172	Total	C	N	O	S	0	0	0
			1409	911	241	253	4			

There are 81 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	28	HIS	GLN	engineered mutation	UNP P80188
A	36	GLU	LEU	engineered mutation	UNP P80188
A	40	SER	ALA	engineered mutation	UNP P80188
A	41	LEU	ILE	engineered mutation	UNP P80188
A	49	ARG	GLN	engineered mutation	UNP P80188
A	70	ARG	LEU	engineered mutation	UNP P80188
A	73	SER	LYS	engineered mutation	UNP P80188
A	77	HIS	ASP	engineered mutation	UNP P80188
A	79	LEU	TRP	engineered mutation	UNP P80188
A	87	SER	CYS	engineered mutation	UNP P80188
A	96	LEU	ASN	engineered mutation	UNP P80188
A	100	LYS	TYR	engineered mutation	UNP P80188
A	103	HIS	LEU	engineered mutation	UNP P80188
A	106	PHE	TYR	engineered mutation	UNP P80188
A	125	THR	LYS	engineered mutation	UNP P80188
A	127	ALA	SER	engineered mutation	UNP P80188
A	134	PHE	LYS	engineered mutation	UNP P80188
A	179	SER	-	expression tag	UNP P80188
A	180	ALA	-	expression tag	UNP P80188
A	181	TRP	-	expression tag	UNP P80188
A	182	SER	-	expression tag	UNP P80188
A	183	HIS	-	expression tag	UNP P80188
A	184	PRO	-	expression tag	UNP P80188

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	185	GLN	-	expression tag	UNP P80188
A	186	PHE	-	expression tag	UNP P80188
A	187	GLU	-	expression tag	UNP P80188
A	188	LYS	-	expression tag	UNP P80188
C	28	HIS	GLN	engineered mutation	UNP P80188
C	36	GLU	LEU	engineered mutation	UNP P80188
C	40	SER	ALA	engineered mutation	UNP P80188
C	41	LEU	ILE	engineered mutation	UNP P80188
C	49	ARG	GLN	engineered mutation	UNP P80188
C	70	ARG	LEU	engineered mutation	UNP P80188
C	73	SER	LYS	engineered mutation	UNP P80188
C	77	HIS	ASP	engineered mutation	UNP P80188
C	79	LEU	TRP	engineered mutation	UNP P80188
C	87	SER	CYS	engineered mutation	UNP P80188
C	96	LEU	ASN	engineered mutation	UNP P80188
C	100	LYS	TYR	engineered mutation	UNP P80188
C	103	HIS	LEU	engineered mutation	UNP P80188
C	106	PHE	TYR	engineered mutation	UNP P80188
C	125	THR	LYS	engineered mutation	UNP P80188
C	127	ALA	SER	engineered mutation	UNP P80188
C	134	PHE	LYS	engineered mutation	UNP P80188
C	179	SER	-	expression tag	UNP P80188
C	180	ALA	-	expression tag	UNP P80188
C	181	TRP	-	expression tag	UNP P80188
C	182	SER	-	expression tag	UNP P80188
C	183	HIS	-	expression tag	UNP P80188
C	184	PRO	-	expression tag	UNP P80188
C	185	GLN	-	expression tag	UNP P80188
C	186	PHE	-	expression tag	UNP P80188
C	187	GLU	-	expression tag	UNP P80188
C	188	LYS	-	expression tag	UNP P80188
E	28	HIS	GLN	engineered mutation	UNP P80188
E	36	GLU	LEU	engineered mutation	UNP P80188
E	40	SER	ALA	engineered mutation	UNP P80188
E	41	LEU	ILE	engineered mutation	UNP P80188
E	49	ARG	GLN	engineered mutation	UNP P80188
E	70	ARG	LEU	engineered mutation	UNP P80188
E	73	SER	LYS	engineered mutation	UNP P80188
E	77	HIS	ASP	engineered mutation	UNP P80188
E	79	LEU	TRP	engineered mutation	UNP P80188
E	87	SER	CYS	engineered mutation	UNP P80188
E	96	LEU	ASN	engineered mutation	UNP P80188

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	100	LYS	TYR	engineered mutation	UNP P80188
E	103	HIS	LEU	engineered mutation	UNP P80188
E	106	PHE	TYR	engineered mutation	UNP P80188
E	125	THR	LYS	engineered mutation	UNP P80188
E	127	ALA	SER	engineered mutation	UNP P80188
E	134	PHE	LYS	engineered mutation	UNP P80188
E	179	SER	-	expression tag	UNP P80188
E	180	ALA	-	expression tag	UNP P80188
E	181	TRP	-	expression tag	UNP P80188
E	182	SER	-	expression tag	UNP P80188
E	183	HIS	-	expression tag	UNP P80188
E	184	PRO	-	expression tag	UNP P80188
E	185	GLN	-	expression tag	UNP P80188
E	186	PHE	-	expression tag	UNP P80188
E	187	GLU	-	expression tag	UNP P80188
E	188	LYS	-	expression tag	UNP P80188

- Molecule 2 is a protein called Fibronectin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	183	Total	C	N	O	S	0	0	0
			1392	883	222	286	1			
2	D	280	Total	C	N	O	S	0	0	0
			2134	1341	345	446	2			
2	F	277	Total	C	N	O	S	0	0	0
			2109	1326	338	443	2			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1449	SER	-	expression tag	UNP P02751
B	1450	ALA	-	expression tag	UNP P02751
B	1451	HIS	-	expression tag	UNP P02751
B	1452	HIS	-	expression tag	UNP P02751
B	1453	HIS	-	expression tag	UNP P02751
B	1454	HIS	-	expression tag	UNP P02751
B	1455	HIS	-	expression tag	UNP P02751
B	1456	HIS	-	expression tag	UNP P02751
D	1449	SER	-	expression tag	UNP P02751
D	1450	ALA	-	expression tag	UNP P02751
D	1451	HIS	-	expression tag	UNP P02751
D	1452	HIS	-	expression tag	UNP P02751

Continued on next page...


Continued from previous page...

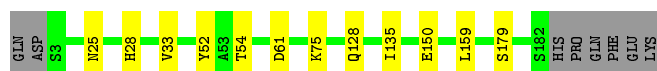
Chain	Residue	Modelled	Actual	Comment	Reference
D	1453	HIS	-	expression tag	UNP P02751
D	1454	HIS	-	expression tag	UNP P02751
D	1455	HIS	-	expression tag	UNP P02751
D	1456	HIS	-	expression tag	UNP P02751
F	1449	SER	-	expression tag	UNP P02751
F	1450	ALA	-	expression tag	UNP P02751
F	1451	HIS	-	expression tag	UNP P02751
F	1452	HIS	-	expression tag	UNP P02751
F	1453	HIS	-	expression tag	UNP P02751
F	1454	HIS	-	expression tag	UNP P02751
F	1455	HIS	-	expression tag	UNP P02751
F	1456	HIS	-	expression tag	UNP P02751

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 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: Neutrophil gelatinase-associated lipocalin

Chain A: 




- Molecule 1: Neutrophil gelatinase-associated lipocalin

Chain C: 



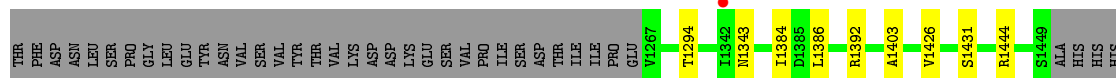
- Molecule 1: Neutrophil gelatinase-associated lipocalin

Chain E: 



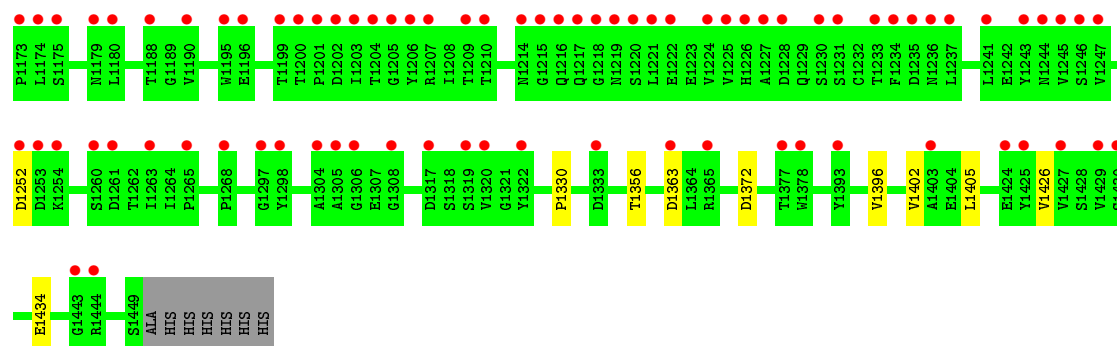
- Molecule 2: Fibronectin

Chain B: 



- Molecule 2: Fibronectin

Chain D: 



4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α , β , γ	135.15Å 135.15Å 203.94Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	35.00 – 3.00 34.92 – 3.00	Depositor EDS
% Data completeness (in resolution range)	99.9 (35.00-3.00) 100.0 (34.92-3.00)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.91 (at 3.00Å)	Xtriage
Refinement program	REFMAC 5.8.0155	Depositor
R, R_{free}	0.208 , 0.241 0.206 , 0.237	Depositor DCC
R_{free} test set	1926 reflections (5.27%)	DCC
Wilson B-factor (Å ²)	70.0	Xtriage
Anisotropy	0.264	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 41.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	9943	wwPDB-VP
Average B, all atoms (Å ²)	99.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.47% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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	0.54	0/1503	0.74	0/2034
1	C	0.54	0/1474	0.73	0/1994
1	E	0.45	0/1447	0.64	0/1957
2	B	0.48	0/1424	0.70	0/1960
2	D	0.44	0/2185	0.67	0/3009
2	F	0.41	0/2158	0.57	0/2972
All	All	0.47	0/10191	0.67	0/13926

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	1463	0	1449	6	0
1	C	1436	0	1425	10	0
1	E	1409	0	1404	3	0
2	B	1392	0	1368	3	0
2	D	2134	0	2068	4	0
2	F	2109	0	2049	2	0
All	All	9943	0	9763	26	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 1.

The worst 5 of 26 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:5:SER:O	1:C:130:ARG:NH2	2.15	0.79
1:C:34:VAL:HG11	1:C:144:LEU:HD13	1.72	0.72
1:A:33:VAL:HG21	1:A:52:TYR:CZ	2.29	0.66
1:C:94:LEU:HD13	1:C:106:PHE:CD1	2.39	0.58
1:C:33:VAL:HG21	1:C:52:TYR:CZ	2.39	0.58

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	178/188 (95%)	169 (95%)	8 (4%)	1 (1%)	28	70
1	C	174/188 (93%)	166 (95%)	7 (4%)	1 (1%)	28	70
1	E	170/188 (90%)	163 (96%)	7 (4%)	0	100	100
2	B	181/284 (64%)	177 (98%)	4 (2%)	0	100	100
2	D	278/284 (98%)	262 (94%)	15 (5%)	1 (0%)	38	78
2	F	275/284 (97%)	259 (94%)	15 (6%)	1 (0%)	38	78
All	All	1256/1416 (89%)	1196 (95%)	56 (4%)	4 (0%)	44	81

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	D	1252	ASP
2	F	1252	ASP
1	C	99	SER
1	A	61	ASP

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	164/172 (95%)	161 (98%)	3 (2%)	64	89
1	C	162/172 (94%)	160 (99%)	2 (1%)	75	93
1	E	158/172 (92%)	155 (98%)	3 (2%)	62	88
2	B	163/256 (64%)	160 (98%)	3 (2%)	64	89
2	D	252/256 (98%)	249 (99%)	3 (1%)	75	93
2	F	250/256 (98%)	245 (98%)	5 (2%)	60	87
All	All	1149/1284 (90%)	1130 (98%)	19 (2%)	66	89

5 of 19 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	1402	VAL
2	D	1431	SER
2	F	1372	ASP
1	C	148	LEU
2	F	1402	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	118	HIS

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, 95th 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 ²)	Q<0.9
1	A	180/188 (95%)	-0.40	0 100 100	39, 56, 106, 129	0
1	C	176/188 (93%)	-0.36	1 (0%) 89 71	42, 62, 112, 166	0
1	E	172/188 (91%)	-0.12	3 (1%) 70 42	58, 96, 157, 222	0
2	B	183/284 (64%)	-0.32	1 (0%) 90 74	41, 67, 101, 142	0
2	D	280/284 (98%)	0.06	17 (6%) 22 8	42, 82, 168, 216	0
2	F	277/284 (97%)	1.35	79 (28%) 1 0	98, 175, 213, 229	0
All	All	1268/1416 (89%)	0.14	101 (7%) 13 5	39, 82, 199, 229	0

The worst 5 of 101 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	1173	PRO	7.7
1	C	99	SER	6.4
2	F	1202	ASP	6.1
2	F	1246	SER	6.0
2	F	1217	GLN	5.9

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.