



# Full wwPDB X-ray Structure Validation Report ⓘ

May 9, 2018 – 11:15 AM EDT

PDB ID : 5NO5  
Title : AbyA5 Wildtype  
Authors : Byrne, M.J.; Race, P.R.  
Deposited on : 2017-04-10  
Resolution : 2.50 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 : rb-20031172  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Refmac : 5.8.0158  
CCP4 : 7.0 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20031172

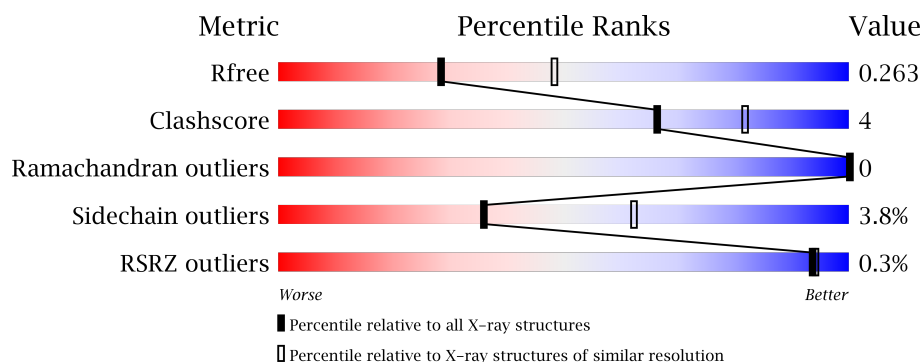
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.50 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))
$R_{free}$	111664	4155 (2.50-2.50)
Clashscore	122126	4827 (2.50-2.50)
Ramachandran outliers	120053	4735 (2.50-2.50)
Sidechain outliers	120020	4737 (2.50-2.50)
RSRZ outliers	108989	4058 (2.50-2.50)

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	374	
1	B	374	

## 2 Entry composition i

There are 2 unique types of molecules in this entry. The entry contains 5399 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 AbyA5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	349	2648	1677	478	490	3	0	0	0
1	B	349	2653	1679	480	491	3	0	0	0

There are 38 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP F4F7F5
A	2	ALA	-	expression tag	UNP F4F7F5
A	3	HIS	-	expression tag	UNP F4F7F5
A	4	HIS	-	expression tag	UNP F4F7F5
A	5	HIS	-	expression tag	UNP F4F7F5
A	6	HIS	-	expression tag	UNP F4F7F5
A	7	HIS	-	expression tag	UNP F4F7F5
A	8	HIS	-	expression tag	UNP F4F7F5
A	9	SER	-	expression tag	UNP F4F7F5
A	10	SER	-	expression tag	UNP F4F7F5
A	11	GLY	-	expression tag	UNP F4F7F5
A	12	LEU	-	expression tag	UNP F4F7F5
A	13	GLU	-	expression tag	UNP F4F7F5
A	14	VAL	-	expression tag	UNP F4F7F5
A	15	LEU	-	expression tag	UNP F4F7F5
A	16	PHE	-	expression tag	UNP F4F7F5
A	17	GLN	-	expression tag	UNP F4F7F5
A	18	GLY	-	expression tag	UNP F4F7F5
A	19	PRO	-	expression tag	UNP F4F7F5
B	1	MET	-	initiating methionine	UNP F4F7F5
B	2	ALA	-	expression tag	UNP F4F7F5
B	3	HIS	-	expression tag	UNP F4F7F5
B	4	HIS	-	expression tag	UNP F4F7F5
B	5	HIS	-	expression tag	UNP F4F7F5
B	6	HIS	-	expression tag	UNP F4F7F5

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	7	HIS	-	expression tag	UNP F4F7F5
B	8	HIS	-	expression tag	UNP F4F7F5
B	9	SER	-	expression tag	UNP F4F7F5
B	10	SER	-	expression tag	UNP F4F7F5
B	11	GLY	-	expression tag	UNP F4F7F5
B	12	LEU	-	expression tag	UNP F4F7F5
B	13	GLU	-	expression tag	UNP F4F7F5
B	14	VAL	-	expression tag	UNP F4F7F5
B	15	LEU	-	expression tag	UNP F4F7F5
B	16	PHE	-	expression tag	UNP F4F7F5
B	17	GLN	-	expression tag	UNP F4F7F5
B	18	GLY	-	expression tag	UNP F4F7F5
B	19	PRO	-	expression tag	UNP F4F7F5

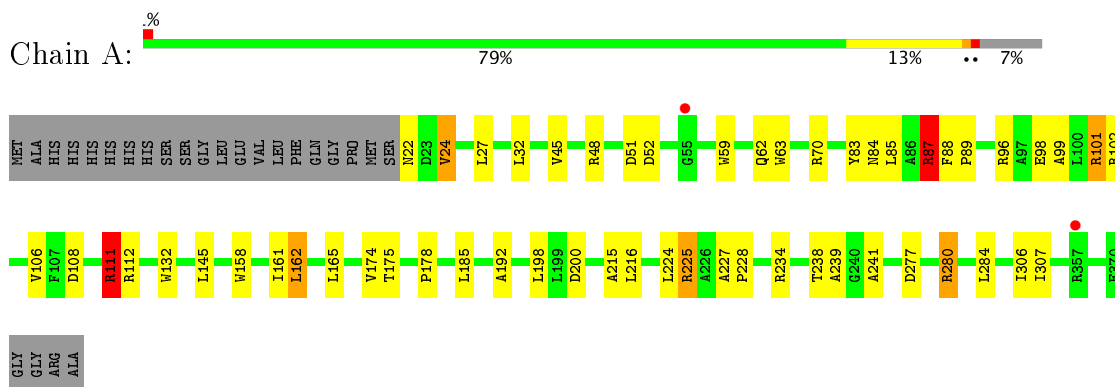
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	45	Total O 45 45	0	0
2	B	53	Total O 53 53	0	0

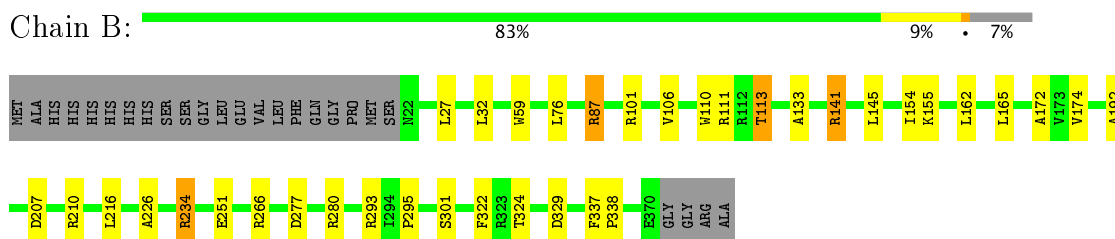
### 3 Residue-property plots [i](#)

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: AbyA5



- Molecule 1: AbyA5



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	69.23Å 88.36Å 131.29Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	73.30 – 2.50 41.87 – 2.50	Depositor EDS
% Data completeness (in resolution range)	96.0 (73.30-2.50) 96.0 (41.87-2.50)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.25 (at 2.51Å)	Xtrriage
Refinement program	REFMAC 5.8.0158	Depositor
R, $R_{free}$	0.208 , 0.261 0.216 , 0.263	Depositor DCC
$R_{free}$ test set	1356 reflections (4.94%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	16.2	Xtrriage
Anisotropy	1.188	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 30.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.35$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	5399	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	16.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 56.95 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 2.5417e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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](#)

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	0.80	2/2712 (0.1%)	0.95	8/3715 (0.2%)
1	B	0.74	0/2717	0.90	6/3721 (0.2%)
All	All	0.77	2/5429 (0.0%)	0.93	14/7436 (0.2%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	158	TRP	CB-CG	-7.04	1.37	1.50
1	A	87	ARG	CZ-NH2	-5.27	1.26	1.33

All (14) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	87	ARG	NE-CZ-NH2	-7.33	116.63	120.30
1	A	280	ARG	NE-CZ-NH1	7.26	123.93	120.30
1	A	101	ARG	NE-CZ-NH1	6.86	123.73	120.30
1	A	87	ARG	NE-CZ-NH1	6.61	123.60	120.30
1	A	200	ASP	CB-CG-OD1	6.13	123.81	118.30
1	B	101	ARG	NE-CZ-NH1	5.88	123.24	120.30
1	B	210	ARG	NE-CZ-NH1	5.88	123.24	120.30
1	A	111	ARG	NE-CZ-NH1	5.86	123.23	120.30
1	A	111	ARG	NE-CZ-NH2	-5.68	117.46	120.30
1	B	293	ARG	NE-CZ-NH1	5.60	123.10	120.30
1	B	234	ARG	NE-CZ-NH1	5.23	122.91	120.30
1	B	141	ARG	NE-CZ-NH2	-5.14	117.73	120.30
1	B	266	ARG	NE-CZ-NH2	-5.11	117.75	120.30
1	A	234	ARG	NE-CZ-NH1	5.04	122.82	120.30

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	2648	0	2593	30	0
1	B	2653	0	2600	16	0
2	A	45	0	0	0	0
2	B	53	0	0	0	0
All	All	5399	0	5193	45	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 4.

All (45) 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:A:84:ASN:OD1	1:A:87:ARG:NH1	1.98	0.94
1:A:238:THR:HG21	1:A:241:ALA:HB2	1.46	0.94
1:A:215:ALA:O	1:A:238:THR:HG23	1.73	0.87
1:B:162:LEU:HD21	1:B:174:VAL:HG23	1.60	0.84
1:B:110:TRP:O	1:B:113:THR:HG22	1.88	0.73
1:A:98:GLU:OE1	1:A:101:ARG:NH1	2.25	0.69
1:A:238:THR:HG22	1:A:239:ALA:H	1.57	0.67
1:A:277:ASP:OD1	1:A:280:ARG:NH1	2.31	0.63
1:A:24:VAL:HG21	1:A:51:ASP:HA	1.80	0.62
1:B:145:LEU:HD23	1:B:165:LEU:HD22	1.81	0.62
1:B:145:LEU:CD2	1:B:165:LEU:HD22	2.32	0.59
1:A:48:ARG:HE	1:A:62:GLN:HE22	1.51	0.58
1:A:238:THR:HG22	1:A:239:ALA:N	2.17	0.57
1:A:111:ARG:HG3	1:A:112:ARG:N	2.19	0.57
1:B:133:ALA:HA	1:B:172:ALA:O	2.05	0.56
1:B:111:ARG:HD2	1:B:111:ARG:C	2.30	0.52
1:A:108:ASP:O	1:A:111:ARG:HG2	2.12	0.50
1:A:98:GLU:O	1:A:102:ARG:HG2	2.11	0.50
1:A:111:ARG:HD2	1:A:132:TRP:CZ3	2.46	0.49
1:A:175:THR:HG21	1:A:198:LEU:HD11	1.95	0.49
1:A:111:ARG:CD	1:A:132:TRP:CZ3	2.97	0.48
1:B:27:LEU:HB3	1:B:59:TRP:CZ3	2.48	0.48
1:B:277:ASP:OD1	1:B:280:ARG:NH1	2.48	0.47

*Continued on next page...*



Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:301:SER:HB3	1:B:329:ASP:O	2.16	0.46
1:A:162:LEU:HD11	1:A:174:VAL:HG23	1.98	0.46
1:A:178:PRO:HB3	1:A:185:LEU:HB2	1.98	0.45
1:A:22:ASN:HB2	1:A:52:ASP:OD2	2.16	0.45
1:A:192:ALA:HB2	1:A:225:ARG:HG3	1.99	0.45
1:A:84:ASN:O	1:A:87:ARG:HD3	2.17	0.44
1:A:145:LEU:CD2	1:A:165:LEU:HD22	2.48	0.43
1:A:162:LEU:CD1	1:A:174:VAL:HG23	2.47	0.43
1:B:141:ARG:HB3	1:B:207:ASP:HB2	2.00	0.43
1:B:322:PHE:CE2	1:B:324:THR:HB	2.54	0.43
1:B:192:ALA:HB1	1:B:226:ALA:HB2	2.01	0.43
1:A:27:LEU:HB3	1:A:59:TRP:CZ3	2.54	0.43
1:A:96:ARG:O	1:A:99:ALA:HB3	2.19	0.43
1:A:70:ARG:HH22	1:B:113:THR:HG23	1.85	0.42
1:B:337:PHE:N	1:B:338:PRO:CD	2.83	0.42
1:A:306:ILE:HG22	1:A:307:ILE:HD13	2.02	0.41
1:A:83:TYR:CE1	1:A:102:ARG:HD2	2.56	0.41
1:A:227:ALA:HB3	1:A:228:PRO:HD3	2.02	0.41
1:B:234:ARG:O	1:B:295:PRO:HD2	2.21	0.41
1:B:87:ARG:HD3	1:B:154:ILE:HD13	2.03	0.41
1:A:88:PHE:HA	1:A:89:PRO:HA	1.88	0.41
1:A:45:VAL:HG11	1:A:63:TRP:CE3	2.57	0.40

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	347/374 (93%)	339 (98%)	8 (2%)	0	100	100
1	B	347/374 (93%)	337 (97%)	10 (3%)	0	100	100
All	All	694/748 (93%)	676 (97%)	18 (3%)	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	264/291 (91%)	252 (96%)	12 (4%)	30	54
1	B	265/291 (91%)	257 (97%)	8 (3%)	44	71
All	All	529/582 (91%)	509 (96%)	20 (4%)	36	62

All (20) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	24	VAL
1	A	32	LEU
1	A	85	LEU
1	A	87	ARG
1	A	106	VAL
1	A	111	ARG
1	A	161	ILE
1	A	162	LEU
1	A	216	LEU
1	A	224	LEU
1	A	225	ARG
1	A	284	LEU
1	B	32	LEU
1	B	76	LEU
1	B	87	ARG
1	B	106	VAL
1	B	113	THR
1	B	155	LYS
1	B	216	LEU
1	B	251	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	62	GLN
1	A	72	GLN
1	B	62	GLN
1	B	278	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, 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	349/374 (93%)	-0.26	2 (0%) 89   90	7, 16, 26, 36	0
1	B	349/374 (93%)	-0.28	0 100   100	7, 16, 26, 34	0
All	All	698/748 (93%)	-0.27	2 (0%) 93   94	7, 16, 26, 36	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	55	GLY	2.6
1	A	357	ARG	2.1

### 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.