



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 4, 2017 – 10:34 PM EDT

PDB ID : 2AUA  
Title : Structure of BC2332: A Protein of Unknown Function from Bacillus cereus  
Authors : Brunzelle, J.S.; Minasov, G.; Shuvalova, L.; Collart, F.R.; Joachimiak, A.; Anderson, W.F.; Midwest Center for Structural Genomics (MCSG)  
Deposited on : unknown  
Resolution : 2.35 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
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030345
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-20030345

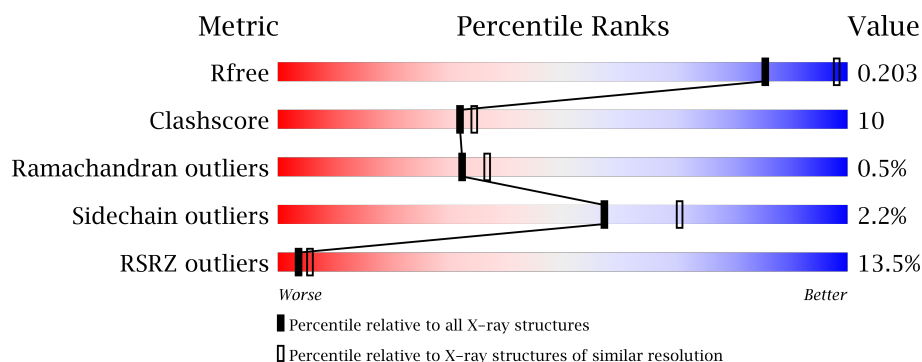
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.35 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	1522 (2.38-2.34)
Clashscore	112137	1626 (2.38-2.34)
Ramachandran outliers	110173	1605 (2.38-2.34)
Sidechain outliers	110143	1606 (2.38-2.34)
RSRZ outliers	101464	1528 (2.38-2.34)

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	224	<div> <div>10%</div> <div> <div></div> <div>74%</div> <div>13%</div> <div>•</div> <div>11%</div> </div> </div>
1	B	224	<div> <div>14%</div> <div> <div></div> <div>72%</div> <div>20%</div> <div>•</div> <div>6%</div> </div> </div>

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 3605 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 hypothetical protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	199	Total	C	N	O	S	Se	0	8	0
			1691	1079	288	319	1	4			
1	B	210	Total	C	N	O	S	Se	0	5	0
			1767	1127	299	335	1	5			

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-23	MSE	-	CLONING ARTIFACT	UNP Q81DM5
A	-22	HIS	-	EXPRESSION TAG	UNP Q81DM5
A	-21	HIS	-	EXPRESSION TAG	UNP Q81DM5
A	-20	HIS	-	EXPRESSION TAG	UNP Q81DM5
A	-19	HIS	-	EXPRESSION TAG	UNP Q81DM5
A	-18	HIS	-	EXPRESSION TAG	UNP Q81DM5
A	-17	HIS	-	EXPRESSION TAG	UNP Q81DM5
A	-16	SER	-	CLONING ARTIFACT	UNP Q81DM5
A	-15	SER	-	CLONING ARTIFACT	UNP Q81DM5
A	-14	GLY	-	CLONING ARTIFACT	UNP Q81DM5
A	-13	VAL	-	CLONING ARTIFACT	UNP Q81DM5
A	-12	ASP	-	CLONING ARTIFACT	UNP Q81DM5
A	-11	LEU	-	CLONING ARTIFACT	UNP Q81DM5
A	-10	GLY	-	CLONING ARTIFACT	UNP Q81DM5
A	-9	THR	-	CLONING ARTIFACT	UNP Q81DM5
A	-8	GLU	-	CLONING ARTIFACT	UNP Q81DM5
A	-7	ASN	-	CLONING ARTIFACT	UNP Q81DM5
A	-6	LEU	-	CLONING ARTIFACT	UNP Q81DM5
A	-5	TYR	-	CLONING ARTIFACT	UNP Q81DM5
A	-4	PHE	-	CLONING ARTIFACT	UNP Q81DM5
A	-3	GLN	-	CLONING ARTIFACT	UNP Q81DM5
A	-2	SER	-	CLONING ARTIFACT	UNP Q81DM5
A	-1	ASN	-	CLONING ARTIFACT	UNP Q81DM5
A	0	ALA	-	CLONING ARTIFACT	UNP Q81DM5
B	-23	MSE	-	CLONING ARTIFACT	UNP Q81DM5

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	-22	HIS	-	EXPRESSION TAG	UNP Q81DM5
B	-21	HIS	-	EXPRESSION TAG	UNP Q81DM5
B	-20	HIS	-	EXPRESSION TAG	UNP Q81DM5
B	-19	HIS	-	EXPRESSION TAG	UNP Q81DM5
B	-18	HIS	-	EXPRESSION TAG	UNP Q81DM5
B	-17	HIS	-	EXPRESSION TAG	UNP Q81DM5
B	-16	SER	-	CLONING ARTIFACT	UNP Q81DM5
B	-15	SER	-	CLONING ARTIFACT	UNP Q81DM5
B	-14	GLY	-	CLONING ARTIFACT	UNP Q81DM5
B	-13	VAL	-	CLONING ARTIFACT	UNP Q81DM5
B	-12	ASP	-	CLONING ARTIFACT	UNP Q81DM5
B	-11	LEU	-	CLONING ARTIFACT	UNP Q81DM5
B	-10	GLY	-	CLONING ARTIFACT	UNP Q81DM5
B	-9	THR	-	CLONING ARTIFACT	UNP Q81DM5
B	-8	GLU	-	CLONING ARTIFACT	UNP Q81DM5
B	-7	ASN	-	CLONING ARTIFACT	UNP Q81DM5
B	-6	LEU	-	CLONING ARTIFACT	UNP Q81DM5
B	-5	TYR	-	CLONING ARTIFACT	UNP Q81DM5
B	-4	PHE	-	CLONING ARTIFACT	UNP Q81DM5
B	-3	GLN	-	CLONING ARTIFACT	UNP Q81DM5
B	-2	SER	-	CLONING ARTIFACT	UNP Q81DM5
B	-1	ASN	-	CLONING ARTIFACT	UNP Q81DM5
B	0	ALA	-	CLONING ARTIFACT	UNP Q81DM5

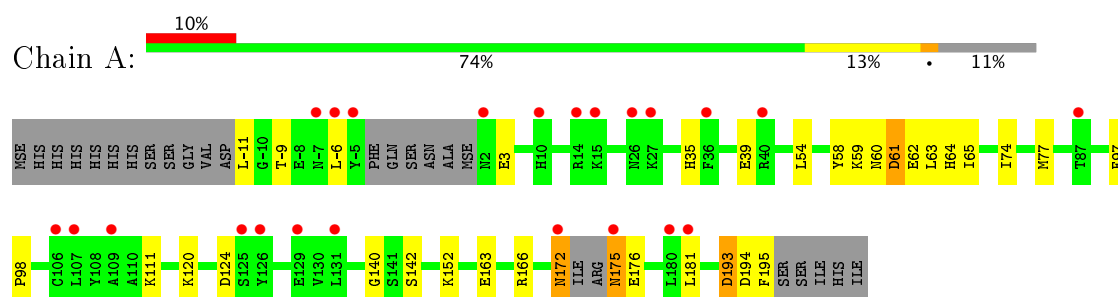
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	62	Total O 63 63	0	1
2	B	84	Total O 84 84	0	0

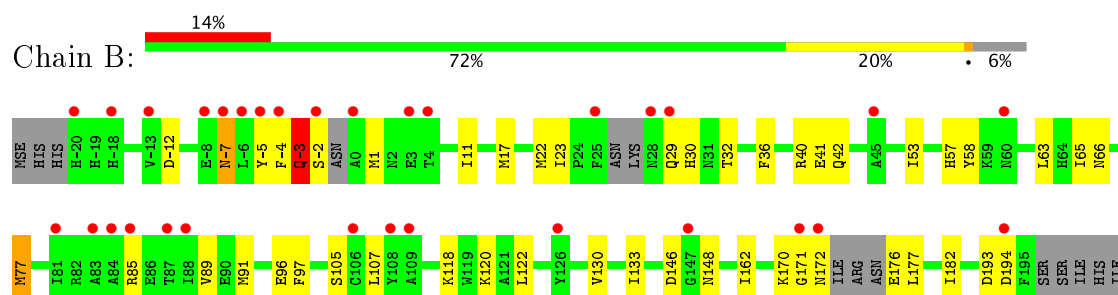
### 3 Residue-property plots [i](#)

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: hypothetical protein



#### • Molecule 1: hypothetical protein



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	72.22Å 76.64Å 107.50Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 2.35 29.61 – 2.33	Depositor EDS
% Data completeness (in resolution range)	99.3 (30.00-2.35) 98.2 (29.61-2.33)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.96 (at 2.34Å)	Xtriage
Refinement program	REFMAC 5.2.0003	Depositor
R, $R_{free}$	0.204 , 0.254 0.205 , 0.203	Depositor DCC
$R_{free}$ test set	1262 reflections (5.25%)	DCC
Wilson B-factor (Å <sup>2</sup> )	45.7	Xtriage
Anisotropy	0.126	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 49.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	3605	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 4.53% 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	0.50	0/1748	0.70	4/2350 (0.2%)
1	B	0.52	0/1813	0.71	4/2435 (0.2%)
All	All	0.51	0/3561	0.71	8/4785 (0.2%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1
1	B	0	1
All	All	0	2

There are no bond length outliers.

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	B	-12	ASP	CB-CG-OD2	6.03	123.72	118.30
1	A	61[A]	ASP	CB-CG-OD2	5.47	123.22	118.30
1	A	61[B]	ASP	CB-CG-OD2	5.47	123.22	118.30
1	A	194	ASP	CB-CG-OD2	5.40	123.16	118.30
1	A	193	ASP	CB-CG-OD2	5.37	123.13	118.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	175	ASN	Peptide
1	B	170	LYS	Peptide

## 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	1691	0	1654	35	0
1	B	1767	0	1703	50	0
2	A	63	0	0	2	0
2	B	84	0	0	3	0
All	All	3605	0	3357	69	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 10.

The worst 5 of 69 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:17:MSE:HE1	1:B:23:ILE:HD11	1.38	1.05
1:B:176:GLU:CG	1:B:177:LEU:H	1.71	1.03
1:A:77:MSE:HE1	1:B:77:MSE:SE	2.13	0.97
1:B:40:ARG:HH11	1:B:40:ARG:HG2	1.29	0.97
1:B:176:GLU:HG3	1:B:177:LEU:H	1.30	0.95

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	201/224 (90%)	192 (96%)	8 (4%)	1 (0%)	32	37
1	B	207/224 (92%)	199 (96%)	7 (3%)	1 (0%)	32	37

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	408/448 (91%)	391 (96%)	15 (4%)	2 (0%)	32 37

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	176	GLU
1	B	-3	GLN

### 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	186/196 (95%)	184 (99%)	2 (1%)	78 87
1	B	193/196 (98%)	187 (97%)	6 (3%)	45 56
All	All	379/392 (97%)	371 (98%)	8 (2%)	57 72

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	-3	GLN
1	B	77	MSE
1	B	1	MSE
1	B	-7	ASN
1	B	-2	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	172	ASN
1	B	-7	ASN
1	B	28	ASN
1	A	95	GLN
1	B	2	ASN

### 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	195/224 (87%)	0.55	23 (11%) 5 9	40, 49, 68, 85	0
1	B	205/224 (91%)	0.70	31 (15%) 3 4	37, 48, 61, 79	0
All	All	400/448 (89%)	0.63	54 (13%) 3 5	37, 49, 63, 85	0

The worst 5 of 54 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	-20	HIS	5.5
1	A	-5	TYR	5.1
1	B	-2	SER	5.0
1	B	28	ASN	4.5
1	B	29	GLN	4.1

### 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.