



# Full wwPDB X-ray Structure Validation Report ⓘ

Jan 17, 2023 – 05:05 PM EST

PDB ID : 1VRA  
Title : Crystal structure of Arginine biosynthesis bifunctional protein argJ (10175521) from *Bacillus halodurans* at 2.00 Å resolution  
Authors : Joint Center for Structural Genomics (JCSG)  
Deposited on : 2005-02-17  
Resolution : 2.00 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.31.2  
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) : 2.31.2

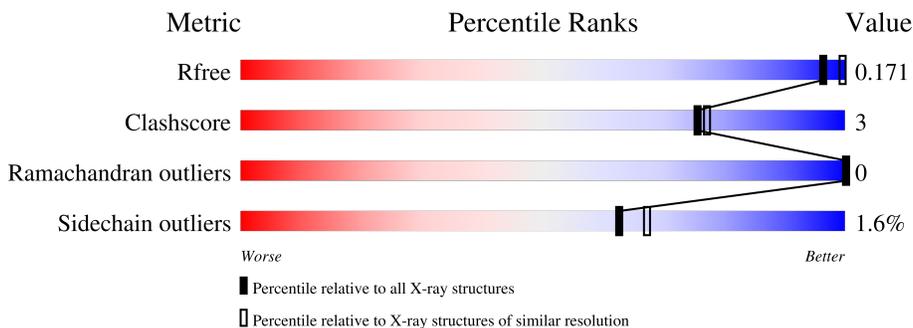
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.00 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))
$R_{free}$	130704	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)

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 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  $\leq 5\%$ .

Mol	Chain	Length	Quality of chain
1	A	208	86% 6% 8%
2	B	215	94% ..

## 2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 3498 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 Arginine biosynthesis bifunctional protein argJ.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	N	O	S				Se
1	A	191	1408	882	242	276	3	5	0	4	0

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-11	MSE	-	expression tag	UNP Q9K8V3
A	-10	GLY	-	expression tag	UNP Q9K8V3
A	-9	SER	-	expression tag	UNP Q9K8V3
A	-8	ASP	-	expression tag	UNP Q9K8V3
A	-7	LYS	-	expression tag	UNP Q9K8V3
A	-6	ILE	-	expression tag	UNP Q9K8V3
A	-5	HIS	-	expression tag	UNP Q9K8V3
A	-4	HIS	-	expression tag	UNP Q9K8V3
A	-3	HIS	-	expression tag	UNP Q9K8V3
A	-2	HIS	-	expression tag	UNP Q9K8V3
A	-1	HIS	-	expression tag	UNP Q9K8V3
A	0	HIS	-	expression tag	UNP Q9K8V3
A	1	MSE	MET	modified residue	UNP Q9K8V3
A	81	MSE	MET	modified residue	UNP Q9K8V3
A	103	MSE	MET	modified residue	UNP Q9K8V3
A	132	MSE	MET	modified residue	UNP Q9K8V3
A	190	MSE	MET	modified residue	UNP Q9K8V3
A	195	MSE	MET	modified residue	UNP Q9K8V3

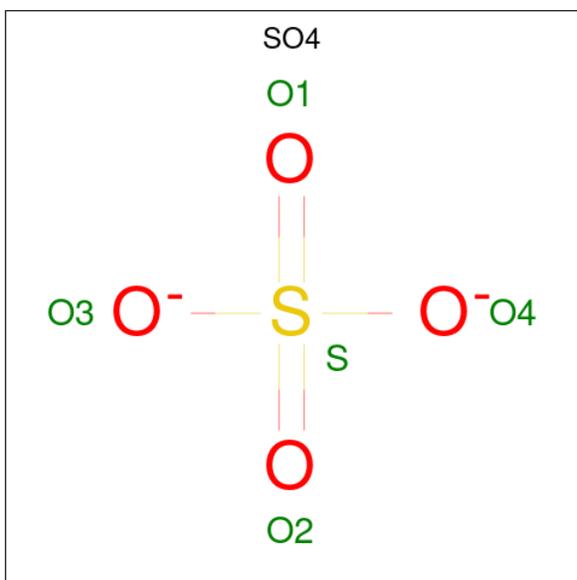
- Molecule 2 is a protein called Arginine biosynthesis bifunctional protein argJ.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	N	O	S				Se
2	B	213	1610	1005	266	330	4	5	0	8	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	198	MSE	MET	modified residue	UNP Q9K8V3
B	239	MSE	MET	modified residue	UNP Q9K8V3
B	243	MSE	MET	modified residue	UNP Q9K8V3
B	252	THR	ALA	SEE REMARK 999	UNP Q9K8V3
B	304	MSE	MET	modified residue	UNP Q9K8V3

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).

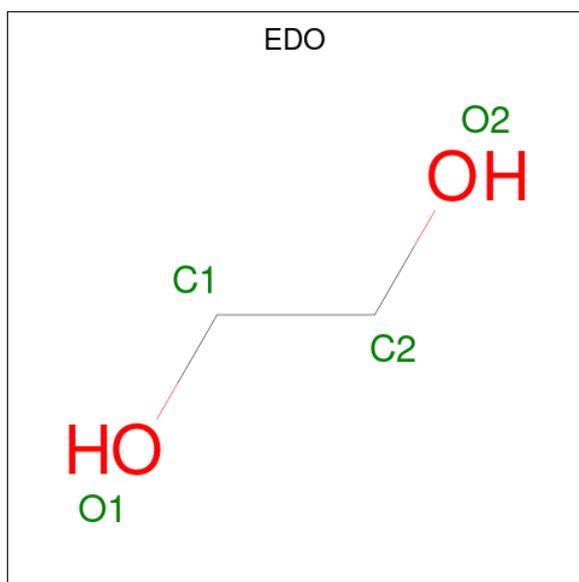


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	O	S	0	0
			5	4	1		
3	A	1	Total	O	S	0	0
			5	4	1		

- Molecule 4 is UNKNOWN LIGAND (three-letter code: UNL) (formula: ).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	O	0	0
			10	10		

- Molecule 5 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C<sub>2</sub>H<sub>6</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total C O 4 2 2	0	0
5	A	1	Total C O 4 2 2	0	0
5	A	1	Total C O 4 2 2	0	0
5	A	1	Total C O 4 2 2	0	0
5	A	1	Total C O 4 2 2	0	0
5	A	1	Total C O 4 2 2	0	0
5	A	1	Total C O 6 3 3	0	1
5	A	1	Total C O 4 2 2	0	0
5	A	1	Total C O 4 2 2	0	0
5	A	1	Total C O 4 2 2	0	0
5	A	1	Total C O 4 2 2	0	0
5	B	1	Total C O 6 3 3	0	1
5	B	1	Total C O 4 2 2	0	0
5	B	1	Total C O 4 2 2	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	B	1	Total	C	O	0	1
			6	3	3		
5	B	1	Total	C	O	0	0
			4	2	2		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	191	Total	O	0	1
			192	192		
6	B	197	Total	O	0	1
			198	198		

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

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. 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. 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: Arginine biosynthesis bifunctional protein argJ

Chain A:  86% 6% 8%



- Molecule 2: Arginine biosynthesis bifunctional protein argJ

Chain B:  94% 4% 2%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	70.66Å 70.66Å 222.29Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.81 – 2.00 29.81 – 2.00	Depositor EDS
% Data completeness (in resolution range)	98.7 (29.81-2.00) 98.7 (29.81-2.00)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	8.30 (at 2.00Å)	Xtrriage
Refinement program	REFMAC 5.2.0005, Xpleo	Depositor
R, $R_{free}$	0.140 , 0.158 0.153 , 0.171	Depositor DCC
$R_{free}$ test set	1935 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	19.5	Xtrriage
Anisotropy	0.146	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 38.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	3498	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	29.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.38% 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

Bond lengths and bond angles in the following residue types are not validated in this section: EDO, SO4, UNL

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	0.57	0/1434	0.65	0/1939
2	B	0.60	0/1665	0.69	0/2259
All	All	0.58	0/3099	0.67	0/4198

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	1408	0	1419	10	0
2	B	1610	0	1543	12	0
3	A	10	0	0	0	0
4	A	10	0	0	0	0
5	A	46	0	72	4	0
5	B	24	0	42	0	0
6	A	192	0	0	3	0
6	B	198	0	0	1	0
All	All	3498	0	3076	19	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 3.

All (19) 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)
5:A:206[B]:EDO:H22	6:A:222:HOH:O	1.83	0.77
5:A:206[C]:EDO:H22	6:A:222:HOH:O	1.86	0.73
2:B:291:VAL:HG23	2:B:310:VAL:HG13	1.75	0.67
2:B:291:VAL:CG2	2:B:310:VAL:HG13	2.27	0.64
2:B:374[A]:GLU:HG2	6:B:457:HOH:O	2.09	0.53
1:A:86:GLY:HA2	5:A:209:EDO:C2	2.41	0.51
1:A:30:HIS:HE1	1:A:34:LYS:O	1.93	0.50
1:A:121:THR:HB	2:B:239[A]:MSE:HE1	1.92	0.50
2:B:294:THR:HG22	2:B:390[B]:THR:HG23	1.93	0.50
1:A:27:VAL:HG23	5:A:204:EDO:O2	2.14	0.48
1:A:65:LEU:CD1	2:B:239[A]:MSE:HE2	2.48	0.44
2:B:291:VAL:HG23	2:B:310:VAL:CG1	2.47	0.43
1:A:65:LEU:HG	2:B:239[A]:MSE:HE2	2.01	0.42
1:A:121:THR:CB	2:B:239[A]:MSE:HE1	2.50	0.42
2:B:295:GLY:HA2	2:B:387:GLY:O	2.20	0.41
1:A:121:THR:HB	2:B:239[B]:MSE:HE1	2.02	0.41
1:A:112:HIS:HD2	6:A:392:HOH:O	2.03	0.41
2:B:352:ILE:HD11	2:B:367:GLU:HG2	2.03	0.41

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	193/208 (93%)	185 (96%)	8 (4%)	0	100	100
2	B	219/215 (102%)	216 (99%)	3 (1%)	0	100	100
All	All	412/423 (97%)	401 (97%)	11 (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	150/158 (95%)	146 (97%)	4 (3%)	44	46
2	B	171/166 (103%)	170 (99%)	1 (1%)	86	90
All	All	321/324 (99%)	316 (98%)	5 (2%)	62	67

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	9	ASN
1	A	61	GLN
1	A	155	ASN
1	A	186	LYS
2	B	310	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	9	ASN
1	A	30	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 monosaccharides in this entry.

## 5.6 Ligand geometry

Of 22 ligands modelled in this entry, 1 is unknown - leaving 21 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
5	EDO	A	206[C]	-	3,3,3	0.56	0	2,2,2	1.12	0
5	EDO	B	17	-	3,3,3	0.49	0	2,2,2	0.34	0
3	SO4	A	198	-	4,4,4	0.20	0	6,6,6	0.43	0
5	EDO	A	207	-	3,3,3	0.42	0	2,2,2	0.40	0
5	EDO	B	14[A]	-	3,3,3	0.39	0	2,2,2	0.43	0
5	EDO	A	202	-	3,3,3	0.40	0	2,2,2	0.62	0
5	EDO	B	7[C]	-	3,3,3	0.43	0	2,2,2	0.42	0
5	EDO	A	208	-	3,3,3	0.45	0	2,2,2	0.34	0
5	EDO	A	201	-	3,3,3	0.48	0	2,2,2	0.31	0
5	EDO	A	206[B]	-	3,3,3	0.43	0	2,2,2	1.44	0
5	EDO	B	10	-	3,3,3	0.53	0	2,2,2	0.53	0
3	SO4	A	197	-	4,4,4	0.07	0	6,6,6	0.38	0
5	EDO	A	205	-	3,3,3	0.36	0	2,2,2	0.55	0
5	EDO	A	210	-	3,3,3	0.35	0	2,2,2	0.52	0
5	EDO	B	7[B]	-	3,3,3	0.56	0	2,2,2	0.25	0
5	EDO	A	203	-	3,3,3	0.53	0	2,2,2	0.11	0
5	EDO	A	204	-	3,3,3	0.48	0	2,2,2	0.14	0
5	EDO	A	200	-	3,3,3	0.66	0	2,2,2	0.27	0
5	EDO	A	209	-	3,3,3	0.56	0	2,2,2	0.32	0
5	EDO	B	11	-	3,3,3	0.37	0	2,2,2	0.12	0
5	EDO	B	14[B]	-	3,3,3	0.43	0	2,2,2	0.15	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
5	EDO	A	206[C]	-	-	1/1/1/1	-
5	EDO	B	17	-	-	1/1/1/1	-
5	EDO	A	207	-	-	0/1/1/1	-
5	EDO	B	14[A]	-	-	0/1/1/1	-

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
5	EDO	A	202	-	-	1/1/1/1	-
5	EDO	B	7[C]	-	-	1/1/1/1	-
5	EDO	A	208	-	-	1/1/1/1	-
5	EDO	A	201	-	-	1/1/1/1	-
5	EDO	A	206[B]	-	-	1/1/1/1	-
5	EDO	B	10	-	-	0/1/1/1	-
5	EDO	A	205	-	-	1/1/1/1	-
5	EDO	A	210	-	-	0/1/1/1	-
5	EDO	B	7[B]	-	-	0/1/1/1	-
5	EDO	A	203	-	-	1/1/1/1	-
5	EDO	A	204	-	-	1/1/1/1	-
5	EDO	A	200	-	-	1/1/1/1	-
5	EDO	A	209	-	-	1/1/1/1	-
5	EDO	B	11	-	-	0/1/1/1	-
5	EDO	B	14[B]	-	-	1/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (13) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	A	201	EDO	O1-C1-C2-O2
5	A	209	EDO	O1-C1-C2-O2
5	A	203	EDO	O1-C1-C2-O2
5	A	206[B]	EDO	O1-C1-C2-O2
5	B	7[C]	EDO	O1-C1-C2-O2
5	A	208	EDO	O1-C1-C2-O2
5	A	205	EDO	O1-C1-C2-O2
5	A	206[C]	EDO	O1-C1-C2-O2
5	B	14[B]	EDO	O1-C1-C2-O2
5	B	17	EDO	O1-C1-C2-O2
5	A	202	EDO	O1-C1-C2-O2
5	A	200	EDO	O1-C1-C2-O2
5	A	204	EDO	O1-C1-C2-O2

There are no ring outliers.

4 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	206[C]	EDO	1	0
5	A	206[B]	EDO	1	0
5	A	204	EDO	1	0
5	A	209	EDO	1	0

## 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

### 6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.