



# Full wwPDB X-ray Structure Validation Report ⓘ

Nov 27, 2023 – 08:02 PM EST

PDB ID : 8GJP  
Title : A1 Int graft: Adenylation domain 1 core construct from teicoplanin biosynthesis, intermediate selection pocket graft  
Authors : Hansen, M.H.; Cryle, M.J.  
Deposited on : 2023-03-16  
Resolution : 2.70 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  
Xtriage (Phenix) : 1.13  
EDS : 2.36  
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) : 2.36

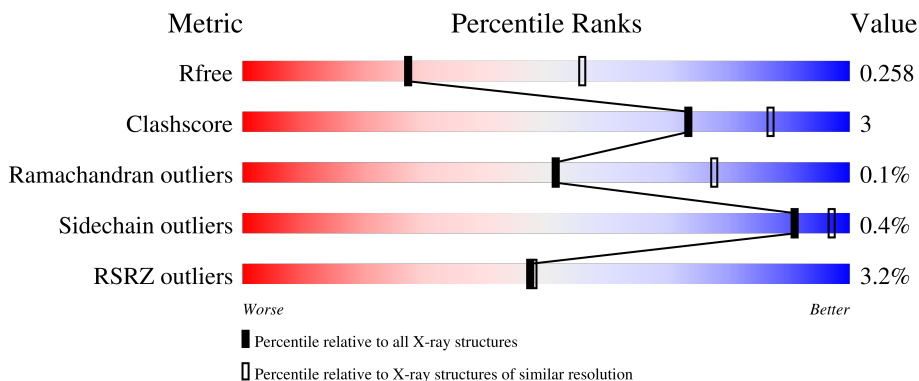
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.70 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	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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  $\leq 5\%$ . 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	399	 3% 88% 8%
1	B	399	 2% 89% 8%
2	C	69	 80% 9% 12%
2	D	69	 12% 87% 6% 7%

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 13578 atoms, of which 6780 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 Non-ribosomal peptide synthetase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	385	Total	C	H	N	O	S	25	12	0
			5835	1826	2933	534	534	8			
1	B	384	Total	C	H	N	O	S	25	14	0
			5832	1819	2936	537	532	8			

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	8	MET	-	initiating methionine	UNP Q70AZ9
A	237	TYR	HIS	engineered mutation	UNP Q70AZ9
A	295	VAL	LEU	engineered mutation	UNP Q70AZ9
A	399	GLY	-	expression tag	UNP Q70AZ9
A	400	SER	-	expression tag	UNP Q70AZ9
A	401	HIS	-	expression tag	UNP Q70AZ9
A	402	HIS	-	expression tag	UNP Q70AZ9
A	403	HIS	-	expression tag	UNP Q70AZ9
A	404	HIS	-	expression tag	UNP Q70AZ9
A	405	HIS	-	expression tag	UNP Q70AZ9
A	406	HIS	-	expression tag	UNP Q70AZ9
B	8	MET	-	initiating methionine	UNP Q70AZ9
B	237	TYR	HIS	engineered mutation	UNP Q70AZ9
B	295	VAL	LEU	engineered mutation	UNP Q70AZ9
B	399	GLY	-	expression tag	UNP Q70AZ9
B	400	SER	-	expression tag	UNP Q70AZ9
B	401	HIS	-	expression tag	UNP Q70AZ9
B	402	HIS	-	expression tag	UNP Q70AZ9
B	403	HIS	-	expression tag	UNP Q70AZ9
B	404	HIS	-	expression tag	UNP Q70AZ9
B	405	HIS	-	expression tag	UNP Q70AZ9
B	406	HIS	-	expression tag	UNP Q70AZ9

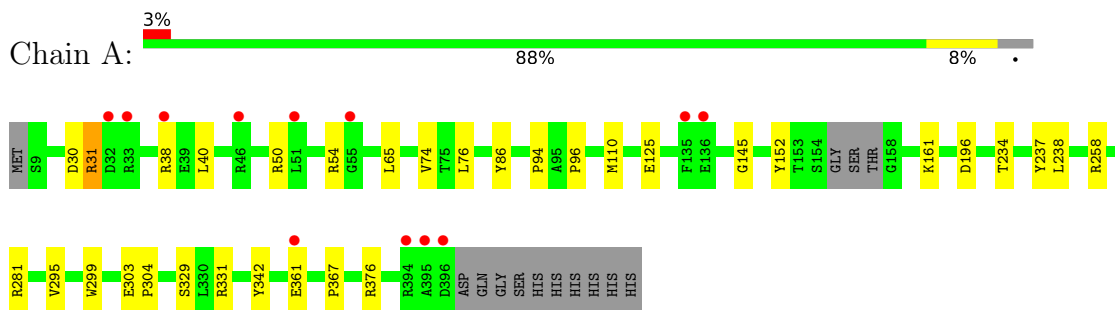
- Molecule 2 is a protein called MbtH-like short polypeptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	H	N	O				S
2	C	61	931	308	442	83	97	1	0	3	0
2	D	64	980	323	469	86	101	1	0	3	0

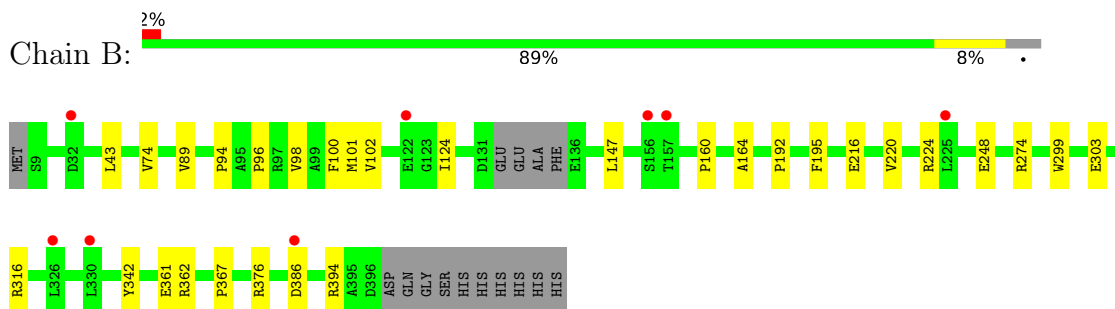
### 3 Residue-property plots [i](#)

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 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.

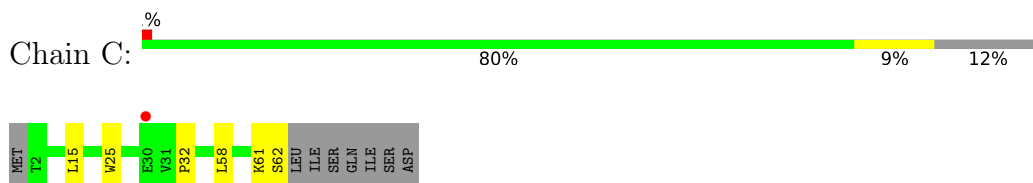
- Molecule 1: Non-ribosomal peptide synthetase



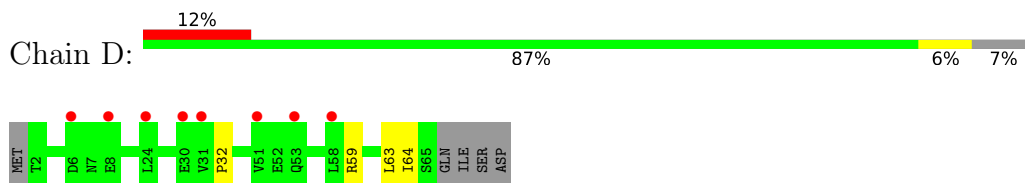
- Molecule 1: Non-ribosomal peptide synthetase



- Molecule 2: MbtH-like short polypeptide



- Molecule 2: MbtH-like short polypeptide



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 2 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	42.75Å 123.42Å 175.79Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	43.95 – 2.70 43.95 – 2.69	Depositor EDS
% Data completeness (in resolution range)	99.5 (43.95-2.70) 99.6 (43.95-2.69)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.11 (at 2.69Å)	Xtrriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, $R_{free}$	0.221 , 0.260 0.221 , 0.258	Depositor DCC
$R_{free}$ test set	1281 reflections (4.82%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	48.8	Xtrriage
Anisotropy	0.808	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.42 , 34.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	13578	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	59.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 14.47% 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 [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	0.31	0/3003	0.66	1/4096 (0.0%)
1	B	0.31	0/3008	0.65	0/4102
2	C	0.26	0/516	0.53	0/707
2	D	0.28	0/538	0.58	0/737
All	All	0.30	0/7065	0.64	1/9642 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	38	ARG	CG-CD-NE	-5.03	101.24	111.80

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	2902	2933	2887	19	1
1	B	2896	2936	2880	19	1
2	C	489	442	434	3	0
2	D	511	469	461	6	0
All	All	6798	6780	6662	44	1

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 (44) 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:394:ARG:HB3	1:B:394:ARG:HH11	1.64	0.63
1:B:43:LEU:HD23	1:B:74:VAL:HG21	1.80	0.63
1:A:54:ARG:NH2	1:A:125:GLU:OE2	2.37	0.55
1:B:367:PRO:HA	2:C:32:PRO:HB3	1.89	0.55
1:A:361:GLU:O	1:A:376:ARG:NE	2.38	0.54
1:B:220:VAL:O	1:B:220:VAL:HG13	2.10	0.52
1:B:342:TYR:CE2	1:B:376:ARG:HG2	2.44	0.52
1:B:100:PHE:CE2	1:B:160:PRO:HG2	2.45	0.51
1:B:361:GLU:O	1:B:376:ARG:NE	2.44	0.51
1:A:196:ASP:O	1:A:295:VAL:HG22	2.11	0.51
1:A:237:TYR:C	1:A:238:LEU:HD12	2.31	0.50
2:D:59:ARG:CZ	2:D:64:ILE:HD12	2.41	0.50
1:B:147[A]:LEU:HD11	1:B:164:ALA:HB1	1.93	0.50
1:B:192:PRO:HG2	1:B:195:PHE:CD2	2.46	0.50
2:D:64:ILE:N	2:D:64:ILE:HD13	2.28	0.49
1:B:248:GLU:O	1:B:274:ARG:NH1	2.48	0.47
1:A:281:ARG:HA	1:A:304:PRO:HB3	1.96	0.47
2:D:59:ARG:NH2	2:D:64:ILE:HD12	2.29	0.47
1:A:94:PRO:HB2	1:A:96:PRO:HD2	1.98	0.46
2:C:61:LYS:O	2:C:62:SER:C	2.54	0.46
1:B:316:ARG:NH2	1:B:386:ASP:O	2.42	0.44
2:C:15:LEU:HD11	2:C:25:TRP:CG	2.52	0.44
1:A:65:LEU:HB2	1:A:110:MET:CE	2.48	0.44
1:A:342:TYR:CE2	1:A:376:ARG:HG2	2.52	0.44
1:B:89:VAL:HG13	1:B:101:MET:HE1	1.98	0.44
1:A:65:LEU:HB2	1:A:110:MET:HE3	2.00	0.43
1:B:94:PRO:HB2	1:B:96:PRO:HD2	2.00	0.43
1:B:394:ARG:HB3	1:B:394:ARG:NH1	2.31	0.43
1:A:54:ARG:NE	1:A:125:GLU:OE2	2.49	0.43
1:A:152:TYR:HA	1:A:161:LYS:O	2.19	0.43
1:A:329:SER:HB2	1:A:331:ARG:HE	1.84	0.42
1:B:98:VAL:HA	1:B:101:MET:HE3	2.01	0.42
1:A:76:LEU:HD22	1:A:86:TYR:CD1	2.54	0.42
1:B:216:GLU:HG3	1:B:224:ARG:NH1	2.35	0.42
1:A:234:THR:HB	1:A:258:ARG:HG3	2.02	0.42
1:B:362[B]:ARG:HA	1:B:376:ARG:HE	1.84	0.42
2:D:59:ARG:HD2	2:D:63:LEU:HD22	2.01	0.41
1:B:147[B]:LEU:HD11	1:B:164:ALA:HB1	2.02	0.41
1:A:30:ASP:O	1:A:31:ARG:C	2.58	0.41
1:A:50:ARG:O	1:A:54:ARG:HG3	2.20	0.41

*Continued on next page...*



Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:367:PRO:HA	2:D:32:PRO:HB3	2.03	0.41
1:B:102:VAL:HG11	1:B:124:ILE:HG13	2.02	0.41
1:A:145:GLY:O	2:D:63:LEU:HD12	2.21	0.41
1:A:40:LEU:HD12	1:A:74:VAL:HG22	2.02	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:303:GLU:OE2	1:B:303:GLU:OE2[4_544]	2.15	0.05

## 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	393/399 (98%)	386 (98%)	6 (2%)	1 (0%)	41	66
1	B	394/399 (99%)	386 (98%)	8 (2%)	0	100	100
2	C	62/69 (90%)	62 (100%)	0	0	100	100
2	D	65/69 (94%)	65 (100%)	0	0	100	100
All	All	914/936 (98%)	899 (98%)	14 (2%)	1 (0%)	51	78

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	31	ARG

### 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	294/295 (100%)	293 (100%)	1 (0%)	92	98
1	B	295/295 (100%)	294 (100%)	1 (0%)	92	98
2	C	54/59 (92%)	53 (98%)	1 (2%)	57	82
2	D	57/59 (97%)	57 (100%)	0	100	100
All	All	700/708 (99%)	697 (100%)	3 (0%)	91	97

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	299	TRP
1	B	299	TRP
2	C	58	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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 [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

### 6.1 Protein, DNA and RNA chains

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	385/399 (96%)	0.42	12 (3%) 49 49	41, 51, 69, 91	0
1	B	384/399 (96%)	0.41	8 (2%) 63 65	40, 52, 71, 105	0
2	C	61/69 (88%)	0.50	1 (1%) 72 74	50, 56, 74, 89	0
2	D	64/69 (92%)	0.76	8 (12%) 3 3	45, 56, 79, 87	0
All	All	894/936 (95%)	0.44	29 (3%) 47 48	40, 52, 72, 105	0

All (29) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	157	THR	3.8
1	B	32	ASP	3.1
2	D	30	GLU	3.1
1	A	32	ASP	2.9
1	A	395	ALA	2.8
1	A	33	ARG	2.7
1	B	330	LEU	2.6
1	A	135	PHE	2.6
1	B	326	LEU	2.5
1	A	394	ARG	2.5
1	A	38	ARG	2.5
1	B	386	ASP	2.5
1	B	122	GLU	2.4
2	D	58	LEU	2.4
2	D	51	VAL	2.4
1	B	225	LEU	2.4
1	A	55	GLY	2.4
1	A	396	ASP	2.4
2	C	30	GLU	2.3
2	D	8	GLU	2.3
2	D	53	GLN	2.3

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	46	ARG	2.3
1	A	136	GLU	2.2
2	D	24	LEU	2.1
1	B	156	SER	2.1
2	D	31	VAL	2.1
2	D	6	ASP	2.1
1	A	51	LEU	2.0
1	A	361	GLU	2.0

## 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 monosaccharides 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.