



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 08:55 AM GMT

PDB ID : 3GJN  
Title : Following evolutionary paths to high affinity and selectivity protein-protein interactions using Colicin7 and Immunity proteins  
Authors : Dym, O.; Tawfik, D.S.  
Deposited on : 2009-03-09  
Resolution : 2.48 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  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
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) : trunk26865

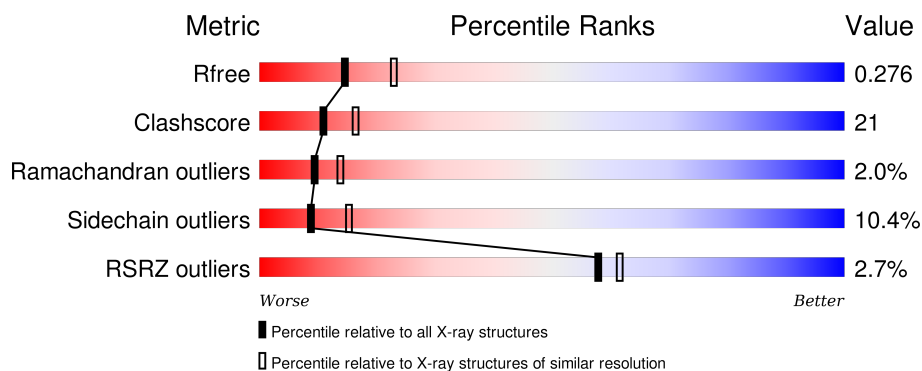
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.48 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))
$R_{free}$	91344	4309 (2.50-2.46)
Clashscore	102246	5050 (2.50-2.46)
Ramachandran outliers	100387	4961 (2.50-2.46)
Sidechain outliers	100360	4963 (2.50-2.46)
RSRZ outliers	91569	4319 (2.50-2.46)

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	86	<div> <div>6%</div> <div>52% 37% 6% 5%</div> </div>
1	D	86	<div> <div>3%</div> <div>63% 27% 6% 5%</div> </div>
2	B	141	<div> <div>2%</div> <div>50% 28% 6% • 16%</div> </div>
2	C	141	<div> <div></div> <div>50% 27% 6% • 15%</div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 3243 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 Colicin-E9 immunity protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	82	Total	C	N	O	S	0	0	0
			647	403	104	138	2			
1	D	82	Total	C	N	O	S	0	0	0
			647	403	104	138	2			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1024	ASP	ASN	ENGINEERED	UNP P13479
A	1026	GLU	ASP	ENGINEERED	UNP P13479
A	1027	ALA	THR	ENGINEERED	UNP P13479
A	1028	THR	SER	ENGINEERED	UNP P13479
A	1034	ASP	VAL	ENGINEERED	UNP P13479
A	1037	ILE	VAL	ENGINEERED	UNP P13479
A	1055	TRP	TYR	ENGINEERED	UNP P13479
D	1024	ASP	ASN	ENGINEERED	UNP P13479
D	1026	GLU	ASP	ENGINEERED	UNP P13479
D	1027	ALA	THR	ENGINEERED	UNP P13479
D	1028	THR	SER	ENGINEERED	UNP P13479
D	1034	ASP	VAL	ENGINEERED	UNP P13479
D	1037	ILE	VAL	ENGINEERED	UNP P13479
D	1055	TRP	TYR	ENGINEERED	UNP P13479

- Molecule 2 is a protein called Colicin-E7.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	119	Total	C	N	O	S	0	0	0
			962	600	182	178	2			
2	C	120	Total	C	N	O	S	0	0	0
			970	605	184	179	2			

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	436	MET	-	EXPRESSION TAG	UNP Q47112
B	437	HIS	-	EXPRESSION TAG	UNP Q47112
B	438	HIS	-	EXPRESSION TAG	UNP Q47112
B	439	HIS	-	EXPRESSION TAG	UNP Q47112
B	440	HIS	-	EXPRESSION TAG	UNP Q47112
B	441	HIS	-	EXPRESSION TAG	UNP Q47112
B	442	HIS	-	EXPRESSION TAG	UNP Q47112
B	443	SER	-	EXPRESSION TAG	UNP Q47112
B	444	MET	-	EXPRESSION TAG	UNP Q47112
B	445	GLY	-	EXPRESSION TAG	UNP Q47112
B	545	ALA	HIS	ENGINEERED	UNP Q47112
C	436	MET	-	EXPRESSION TAG	UNP Q47112
C	437	HIS	-	EXPRESSION TAG	UNP Q47112
C	438	HIS	-	EXPRESSION TAG	UNP Q47112
C	439	HIS	-	EXPRESSION TAG	UNP Q47112
C	440	HIS	-	EXPRESSION TAG	UNP Q47112
C	441	HIS	-	EXPRESSION TAG	UNP Q47112
C	442	HIS	-	EXPRESSION TAG	UNP Q47112
C	443	SER	-	EXPRESSION TAG	UNP Q47112
C	444	MET	-	EXPRESSION TAG	UNP Q47112
C	445	GLY	-	EXPRESSION TAG	UNP Q47112
C	545	ALA	HIS	ENGINEERED	UNP Q47112

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total Zn 1 1	0	0
3	C	1	Total Zn 1 1	0	0

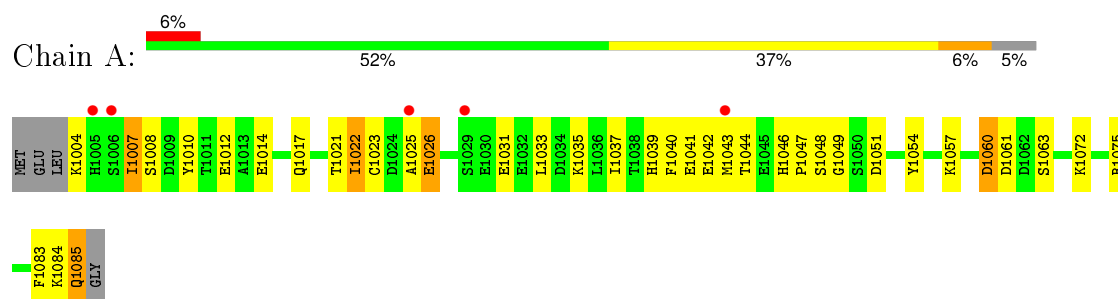
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total O 1 1	0	0
4	B	7	Total O 7 7	0	0
4	C	7	Total O 7 7	0	0

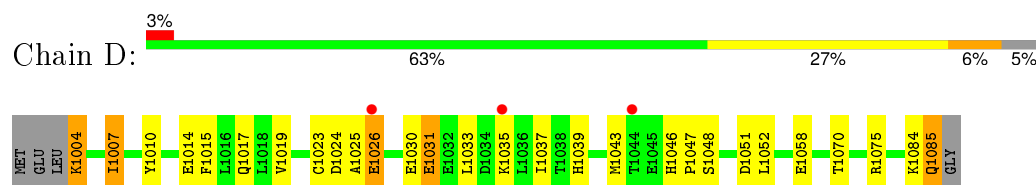
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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 ( $\text{RSRZ} > 2$ ). 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.

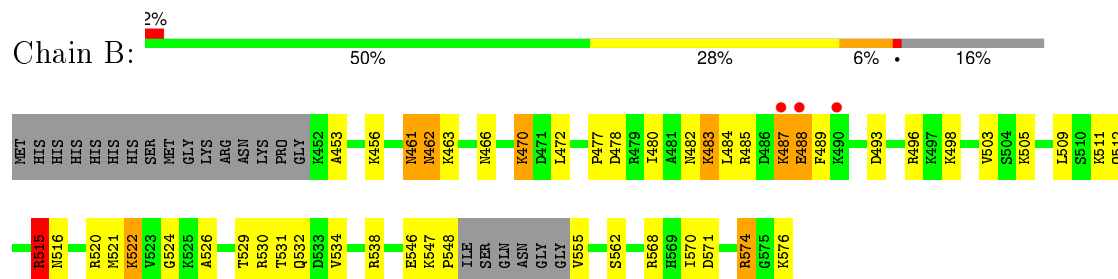
- Molecule 1: Colicin-E9 immunity protein



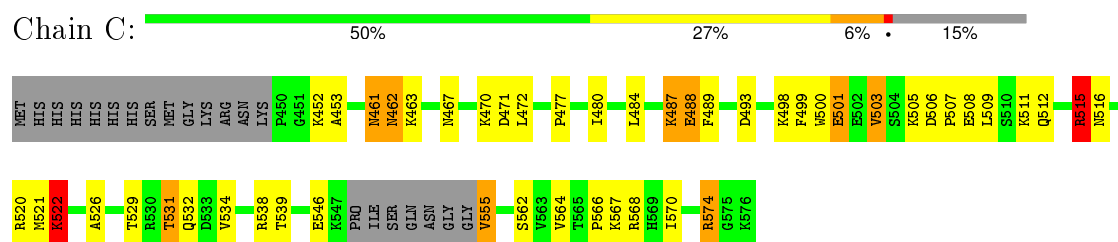
- Molecule 1: Colicin-E9 immunity protein



- Molecule 2: Colicin-E7



- Molecule 2: Colicin-E7



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	59.65Å 53.71Å 79.33Å 90.00° 105.59° 90.00°	Depositor
Resolution (Å)	43.94 – 2.48 43.94 – 2.48	Depositor EDS
% Data completeness (in resolution range)	91.9 (43.94-2.48) 92.1 (43.94-2.48)	Depositor EDS
$R_{merge}$	0.15	Depositor
$R_{sym}$	0.14	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.66 (at 2.48Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.242 , 0.276 0.248 , 0.276	Depositor DCC
$R_{free}$ test set	814 reflections (5.37%)	DCC
Wilson B-factor (Å <sup>2</sup> )	41.6	Xtriage
Anisotropy	0.054	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.40 , 51.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	1 of 15974 reflections (0.006%)	Xtriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	3243	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	39.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 11.51% 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, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	1.30	2/661 (0.3%)	1.17	0/895
1	D	1.34	1/661 (0.2%)	1.14	1/895 (0.1%)
2	B	1.28	1/981 (0.1%)	1.25	9/1311 (0.7%)
2	C	1.37	4/989 (0.4%)	1.25	7/1319 (0.5%)
All	All	1.32	8/3292 (0.2%)	1.21	17/4420 (0.4%)

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
2	B	0	1
2	C	0	1
All	All	0	2

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	1054	TYR	CD1-CE1	6.60	1.49	1.39
2	C	555	VAL	CB-CG2	6.00	1.65	1.52
2	C	503	VAL	CB-CG2	5.95	1.65	1.52
2	C	501	GLU	CG-CD	5.73	1.60	1.51
2	C	500	TRP	CE3-CZ3	5.31	1.47	1.38

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	515	ARG	NE-CZ-NH1	8.44	124.52	120.30
2	B	485	ARG	NE-CZ-NH2	6.75	123.68	120.30

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	483	LYS	CD-CE-NZ	6.65	126.98	111.70
2	B	568	ARG	NE-CZ-NH1	6.40	123.50	120.30
2	C	567	LYS	CD-CE-NZ	-6.24	97.36	111.70

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	555	VAL	Peptide
2	C	555	VAL	Peptide

## 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	647	0	601	31	0
1	D	647	0	601	30	0
2	B	962	0	963	40	0
2	C	970	0	978	39	0
3	B	1	0	0	0	0
3	C	1	0	0	0	0
4	A	1	0	0	1	0
4	B	7	0	0	6	0
4	C	7	0	0	1	0
All	All	3243	0	3143	134	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 21.

The worst 5 of 134 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)
2:B:520:ARG:NH1	2:B:526:ALA:O	1.77	1.17
2:B:470:LYS:HE3	4:B:13:HOH:O	1.45	1.16
1:A:1025:ALA:O	1:A:1026:GLU:HB2	1.49	1.06

*Continued on next page...*



*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:1025:ALA:O	1:D:1026:GLU:HB2	1.57	1.03
2:B:574:ARG:HG3	2:B:574:ARG:HH11	1.22	0.99

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	80/86 (93%)	68 (85%)	9 (11%)	3 (4%)	4	4
1	D	80/86 (93%)	72 (90%)	7 (9%)	1 (1%)	15	24
2	B	115/141 (82%)	105 (91%)	8 (7%)	2 (2%)	11	18
2	C	116/141 (82%)	102 (88%)	12 (10%)	2 (2%)	11	18
All	All	391/454 (86%)	347 (89%)	36 (9%)	8 (2%)	9	14

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	1026	GLU
1	A	1022	ILE
1	A	1023	CYS
2	B	522	LYS
1	D	1026	GLU

### 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	72/75 (96%)	65 (90%)	7 (10%)	10	18
1	D	72/75 (96%)	65 (90%)	7 (10%)	10	18
2	B	105/125 (84%)	94 (90%)	11 (10%)	8	15
2	C	106/125 (85%)	94 (89%)	12 (11%)	7	12
All	All	355/400 (89%)	318 (90%)	37 (10%)	9	15

5 of 37 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	532	GLN
2	C	467	ASN
1	D	1035	LYS
2	B	574	ARG
2	C	461	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
2	C	461	ASN
2	C	462	ASN
1	D	1039	HIS
2	B	516	ASN
2	C	532	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.7 Other polymers

There are no such residues in this entry.

## 5.8 Polymer linkage issues

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	82/86 (95%)	0.75	5 (6%) 25 27	23, 43, 53, 59	0
1	D	82/86 (95%)	0.58	3 (3%) 45 50	23, 43, 53, 59	0
2	B	119/141 (84%)	0.47	3 (2%) 61 64	23, 37, 53, 63	0
2	C	120/141 (85%)	0.31	0 100 100	23, 37, 51, 62	0
All	All	403/454 (88%)	0.50	11 (2%) 58 61	23, 40, 53, 63	0

The worst 5 of 11 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1006	SER	3.0
1	D	1044	THR	2.8
1	A	1005	HIS	2.7
2	B	490	LYS	2.7
1	A	1029	SER	2.6

### 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](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors

of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	ZN	B	600	1/1	0.99	0.16	-	38,38,38,38	0
3	ZN	C	600	1/1	0.98	0.16	-	44,44,44,44	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.