



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 10, 2021 – 01:40 PM EDT

PDB ID : 3GKL
Title : Following evolutionary paths to high affinity and selectivity protein-protein interactions using Colicin7 and Immunity proteins
Authors : Dym, O.; Tawfik, D.S.; Israel Structural Proteomics Center (ISPC)
Deposited on : 2009-03-11
Resolution : 2.20 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

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	:	2.23.2
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)	:	2.23.2

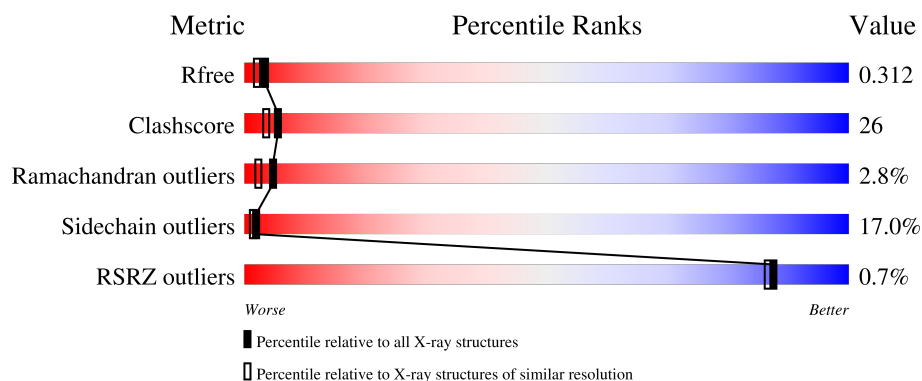
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.20 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	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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 ≥ 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	141	
1	B	141	
2	C	86	
2	D	86	

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 3223 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	120	Total	C	N	O	S	0	0	0
			971	605	185	179	2			
1	B	120	Total	C	N	O	S	0	0	0
			971	605	185	179	2			

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	436	MET	-	expression tag	UNP Q47112
A	437	HIS	-	expression tag	UNP Q47112
A	438	HIS	-	expression tag	UNP Q47112
A	439	HIS	-	expression tag	UNP Q47112
A	440	HIS	-	expression tag	UNP Q47112
A	441	HIS	-	expression tag	UNP Q47112
A	442	HIS	-	expression tag	UNP Q47112
A	443	SER	-	expression tag	UNP Q47112
A	444	MET	-	expression tag	UNP Q47112
A	445	GLY	-	expression tag	UNP Q47112
A	545	ALA	HIS	engineered mutation	UNP Q47112
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 mutation	UNP Q47112

- Molecule 2 is a protein called Colicin-E7.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	82	Total	C	N	O	S	0	0	0
			637	395	102	138	2			
2	D	82	Total	C	N	O	S	0	0	0
			632	392	101	137	2			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	1020	ALA	THR	engineered mutation	UNP P13479
C	1024	ASP	ASN	engineered mutation	UNP P13479
C	1027	ALA	THR	engineered mutation	UNP P13479
C	1028	THR	SER	engineered mutation	UNP P13479
C	1034	ASP	VAL	engineered mutation	UNP P13479
C	1037	ILE	VAL	engineered mutation	UNP P13479
C	1041	GLY	GLU	engineered mutation	UNP P13479
C	1057	GLU	LYS	engineered mutation	UNP P13479
D	1020	ALA	THR	engineered mutation	UNP P13479
D	1024	ASP	ASN	engineered mutation	UNP P13479
D	1027	ALA	THR	engineered mutation	UNP P13479
D	1028	THR	SER	engineered mutation	UNP P13479
D	1034	ASP	VAL	engineered mutation	UNP P13479
D	1037	ILE	VAL	engineered mutation	UNP P13479
D	1041	GLY	GLU	engineered mutation	UNP P13479
D	1057	GLU	LYS	engineered mutation	UNP P13479

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

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

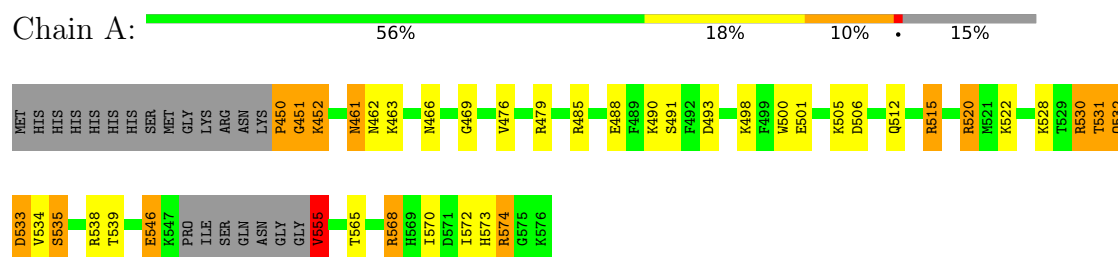
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	5	Total	O	0	0
			5	5		
4	B	3	Total	O	0	0
			3	3		
4	D	2	Total	O	0	0
			2	2		

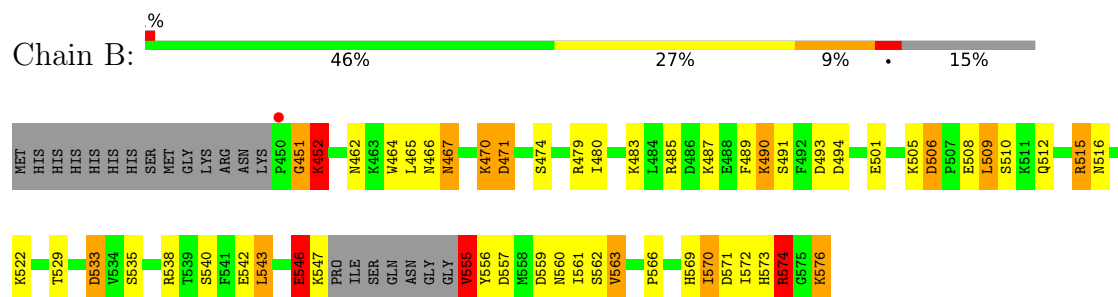
3 Residue-property plots [i](#)

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.

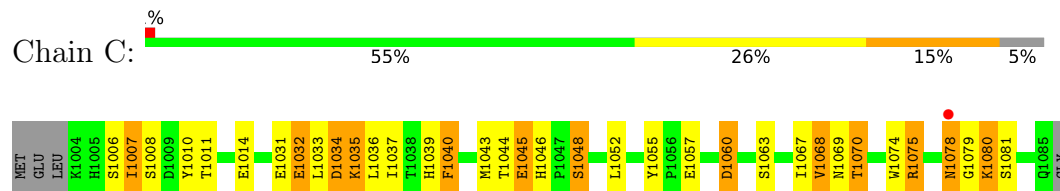
• Molecule 1: Colicin-E9 immunity protein



• 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 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	54.16Å 67.37Å 123.23Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.20 45.47 – 2.20	Depositor EDS
% Data completeness (in resolution range)	95.4 (50.00-2.20) 95.4 (45.47-2.20)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	0.12	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.34 (at 2.20Å)	Xtriage
Refinement program	REFMAC 5.4.0067	Depositor
R, R_{free}	0.240 , 0.277 0.267 , 0.312	Depositor DCC
R_{free} test set	1147 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å ²)	39.8	Xtriage
Anisotropy	0.053	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 59.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.45$, $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	3223	wwPDB-VP
Average B, all atoms (Å ²)	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 34.03 % 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 7.2747e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹Intensities estimated from amplitudes.

²Theoretical 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](#)

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	1.59	2/990 (0.2%)	1.14	4/1321 (0.3%)
1	B	1.63	3/990 (0.3%)	1.09	8/1321 (0.6%)
2	C	1.54	0/650	0.93	1/880 (0.1%)
2	D	1.61	4/644 (0.6%)	0.92	0/871
All	All	1.60	9/3274 (0.3%)	1.04	13/4393 (0.3%)

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	2
1	B	0	2
2	C	0	2
2	D	2	1
All	All	2	7

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	546	GLU	C-N	6.96	1.50	1.34
2	D	1057	GLU	CG-CD	-5.83	1.43	1.51
2	D	1071	VAL	C-N	5.54	1.46	1.34
2	D	1055	TYR	CD1-CE1	-5.51	1.31	1.39
1	A	500	TRP	CD2-CE2	-5.48	1.34	1.41

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	574	ARG	NE-CZ-NH1	8.12	124.36	120.30
1	B	546	GLU	O-C-N	-5.84	113.36	122.70
1	A	574	ARG	CG-CD-NE	5.70	123.76	111.80
1	A	520	ARG	CG-CD-NE	-5.69	99.85	111.80
1	B	451	GLY	N-CA-C	-5.65	98.98	113.10

All (2) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
2	D	1004	LYS	CA
2	D	1027	ALA	CA

5 of 7 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	450	PRO	Peptide
1	A	555	VAL	Peptide
1	B	546	GLU	Mainchain
1	B	555	VAL	Peptide
2	C	1031	GLU	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	971	0	979	36	0
1	B	971	0	979	53	1
2	C	637	0	586	37	0
2	D	632	0	575	40	1
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	5	0	0	0	0
4	B	3	0	0	0	0
4	D	2	0	0	0	0
All	All	3223	0	3119	164	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 26.

The worst 5 of 164 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:D:1046:HIS:CD2	2:D:1047:PRO:HD2	1.36	1.55
1:A:538:ARG:NH1	1:A:570:ILE:HD11	1.43	1.28
2:C:1007:ILE:H	2:C:1007:ILE:CD1	1.50	1.18
2:D:1046:HIS:CD2	2:D:1047:PRO:CD	2.30	1.13
1:A:461:ASN:ND2	1:A:463:LYS:H	1.48	1.10

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:B:576:LYS:NZ	2:D:1058:GLU:OE2[4_456]	2.17	0.03

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	116/141 (82%)	102 (88%)	13 (11%)	1 (1%)	17	16
1	B	116/141 (82%)	102 (88%)	10 (9%)	4 (3%)	3	1
2	C	80/86 (93%)	70 (88%)	6 (8%)	4 (5%)	2	0
2	D	78/86 (91%)	68 (87%)	8 (10%)	2 (3%)	5	3
All	All	390/454 (86%)	342 (88%)	37 (10%)	11 (3%)	5	2

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	451	GLY
2	C	1032	GLU
2	C	1079	GLY
2	C	1045	GLU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	510	SER

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	107/125 (86%)	85 (79%)	22 (21%)	1	1
1	B	107/125 (86%)	91 (85%)	16 (15%)	3	2
2	C	70/73 (96%)	58 (83%)	12 (17%)	2	1
2	D	69/73 (94%)	59 (86%)	10 (14%)	3	2
All	All	353/396 (89%)	293 (83%)	60 (17%)	2	1

5 of 60 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	493	ASP
2	D	1031	GLU
1	B	555	VAL
2	D	1028	THR
2	D	1084	LYS

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

Mol	Chain	Res	Type
1	B	560	ASN
2	C	1039	HIS
2	D	1046	HIS
2	C	1046	HIS
1	A	532	GLN

5.3.3 RNA ⓘ

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](#)

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 [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, 95th 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 ²)	Q<0.9
1	A	120/141 (85%)	-0.27	0 100 100	15, 40, 56, 65	0
1	B	120/141 (85%)	-0.16	1 (0%) 86 85	17, 41, 57, 67	0
2	C	82/86 (95%)	-0.15	1 (1%) 79 77	15, 48, 58, 65	0
2	D	82/86 (95%)	-0.11	1 (1%) 79 77	18, 48, 57, 64	0
All	All	404/454 (88%)	-0.18	3 (0%) 87 86	15, 44, 58, 67	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	450	PRO	3.8
2	D	1007	ILE	2.3
2	C	1078	ASN	2.3

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](#)

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. The B-factors column lists the minimum, median, 95th 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	B-factors(\AA^2)	Q<0.9
3	ZN	B	600	1/1	0.27	0.16	47,47,47,47	0
3	ZN	A	600	1/1	0.95	0.16	45,45,45,45	0

6.5 Other polymers [i](#)

There are no such residues in this entry.