WORLDWIDE wwPDB PROTEIN DATA BANK

Full wwPDB NMR Structure Validation Report (i)

Nov 7, 2023 – 05:04 pm GMT

PDB ID	:	7ZED
BMRB ID	:	34720
Title	:	Solution structure of thanatin-like derivative 7 in complex with E.coli LptA mutant Q62L
Authors	:	Moehle, K.; Zerbe, O.
Deposited on	:	2022-03-31

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

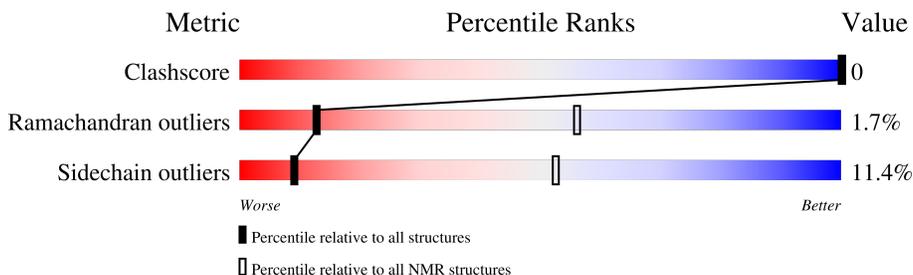|		
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
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:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 83%.

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	Percentile Ranks	Value
Clashscore		0
Ramachandran outliers		1.7%
Sidechain outliers		11.4%

Worse   *Better*

▮ Percentile relative to all structures
▯ Percentile relative to all NMR structures

Metric	Whole archive (#Entries)	NMR archive (#Entries)
Clashscore	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428

The table below summarises the geometric issues observed across the polymeric chains and their fit to the experimental data. The red, orange, yellow and green segments indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A cyan segment indicates the fraction of residues that are not part of the well-defined cores, and 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 <=5%

Mol	Chain	Length	Quality of chain
1	A	118	86% • 12%
2	B	16	50% 12% 38%

WORLDWIDE
wwPDB
PROTEIN DATA BANK

2 Ensemble composition and analysis ⓘ

This entry contains 20 models. Model 12 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *target function.*

The following residues are included in the computation of the global validation metrics.

Well-defined (core) protein residues			
Well-defined core	Residue range (total)	Backbone RMSD (Å)	Medoid model
1	A:36-A:76, A:83-A:145, B:208-B:210, B:212-B:213, B:215-B:215, B:217-B:218, B:220-B:221 (114)	1.52	12

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

The models can be grouped into 2 clusters and 15 single-model clusters were found.

Cluster number	Models
1	11, 12, 13
2	1, 6
Single-model clusters	2; 3; 4; 5; 7; 8; 9; 10; 14; 15; 16; 17; 18; 19; 20

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2075 atoms, of which 1034 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Lipopolysaccharide export system protein LptA.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	118	1789	565	887	151	183	3	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	62	LEU	GLN	engineered mutation	UNP P0ADV1

- Molecule 2 is a protein called Thanatin-like derivative.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	16	286	85	147	29	23	2	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	209	THR	ILE	engineered mutation	UNP P55788
B	211	LE1	CYS	modified residue	UNP P55788
B	214	DAB	ARG	modified residue	UNP P55788
B	216	4FO	GLY	modified residue	UNP P55788
B	219	DAB	GLN	modified residue	UNP P55788
B	221	TYR	MET	engineered mutation	UNP P55788

4 Residue-property plots ⓘ

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-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 outliers are shown as green connectors. Residues which are classified as ill-defined in the NMR ensemble, are shown in cyan with an underline colour-coded according to the previous scheme. Residues which were present in the experimental sample, but not modelled in the final structure are shown in grey.

- Molecule 1: Lipopolysaccharide export system protein LptA

Chain A:

Chain B:

- Molecule 2: Thanatin-like derivative
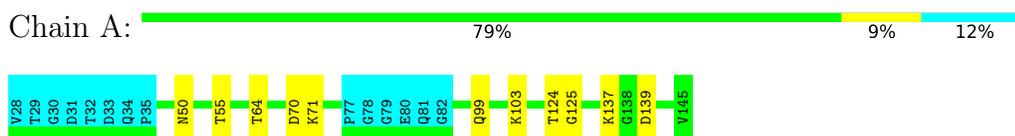
Chain B:
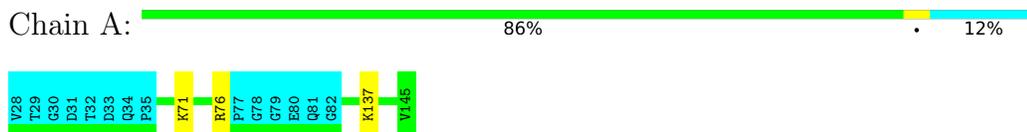
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

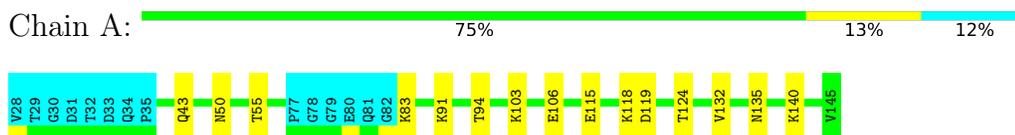- Molecule 1: Lipopolysaccharide export system protein LptA

Chain A:

- Molecule 2: Thanatin-like derivative

Chain B:

4.2.2 Score per residue for model 2

• Molecule 1: Lipopolysaccharide export system protein LptA

Chain A:

• Molecule 2: Thanatin-like derivative

Chain B:

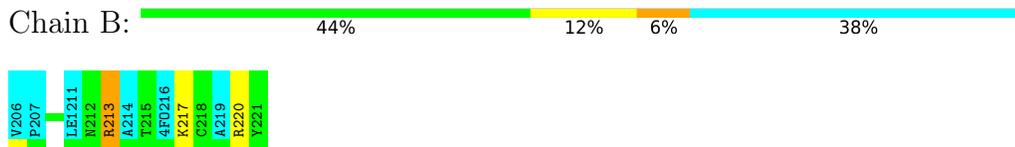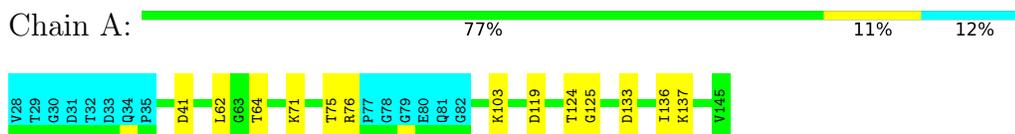### 4.2.3 Score per residue for model 3

• Molecule 1: Lipopolysaccharide export system protein LptA

Chain A:

• Molecule 2: Thanatin-like derivative

Chain B:

4.2.4 Score per residue for model 4

• Molecule 1: Lipopolysaccharide export system protein LptA

Chain A:

• Molecule 2: Thanatin-like derivative

Chain B:

4.2.5 Score per residue for model 5

• Molecule 1: Lipopolysaccharide export system protein LptA

Chain A:
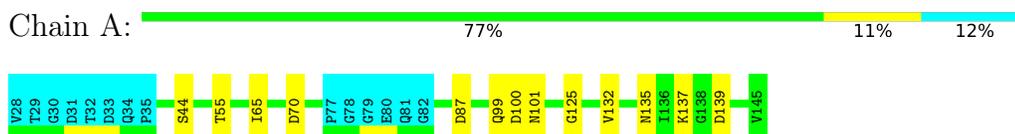

• Molecule 2: Thanatin-like derivative

Chain B:

4.2.6 Score per residue for model 6

• Molecule 1: Lipopolysaccharide export system protein LptA

Chain A:

• Molecule 2: Thanatin-like derivative

Chain B:

4.2.7 Score per residue for model 7

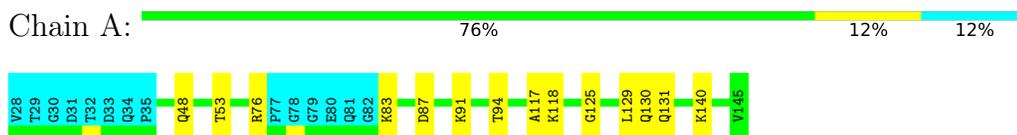• Molecule 1: Lipopolysaccharide export system protein LptA

Chain A:

• Molecule 2: Thanatin-like derivative

Chain B:

4.2.8 Score per residue for model 8

• Molecule 1: Lipopolysaccharide export system protein LptA

Chain A:
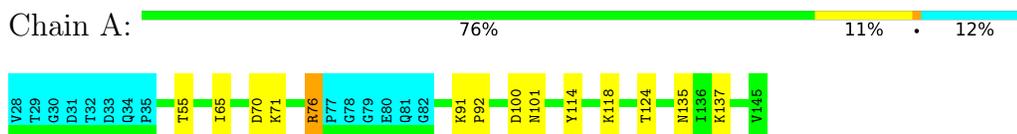

• Molecule 2: Thanatin-like derivative

Chain B:

4.2.9 Score per residue for model 9

• Molecule 1: Lipopolysaccharide export system protein LptA

Chain A:

• Molecule 2: Thanatin-like derivative

Chain B:

4.2.10 Score per residue for model 10

• Molecule 1: Lipopolysaccharide export system protein LptA
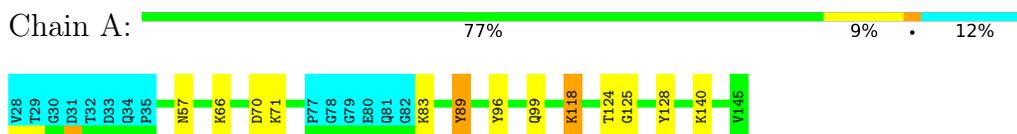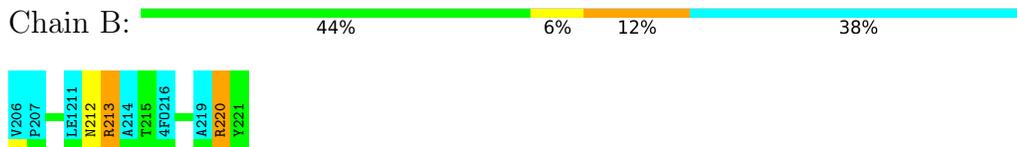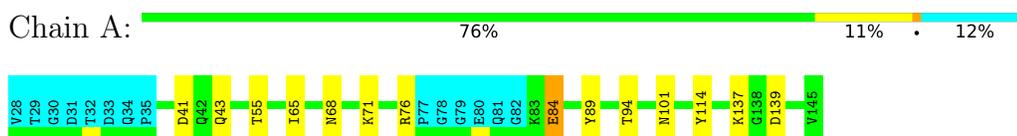
Chain A:

• Molecule 2: Thanatin-like derivative

Chain B:

4.2.11   Score per residue for model 11

• Molecule 1: Lipopolysaccharide export system protein LptA

Chain A:

82%   5% ·   12%

V28 T29 G30 D31 T32 D33 Q34 P35 D70 K71 R76 P77 G78 G79 E80 Q81 G82 M112 T124 G125 K140 V145

• Molecule 2: Thanatin-like derivative

Chain B:

50%   6%   6%   38%

V206 P207 LE1211 N212 R213 A214 T215 4F0216 A219 R220 Y221

4.2.12   Score per residue for model 12 (medoid)

• Molecule 1: Lipopolysaccharide export system protein LptA

Chain A:

75%   13%   12%

V28 T29 G30 D31 T32 D33 Q34 P35 E39 L45 T53 L62 N68 A69 D70 K71 R76 P77 G78 G79 E80 Q81 G82 K91 E106 E115 T124 L129 N135 I136 V145

• Molecule 2: Thanatin-like derivative

Chain B:

44%   19%   38%

V206 P207 LE1211 N212 R213 A214 T215 4F0216 K217 C218 A219 R220 Y221

4.2.13   Score per residue for model 13

• Molecule 1: Lipopolysaccharide export system protein LptA

Chain A:

79%   8%   ·   12%

V28 T29 G30 D31 T32 D33 Q34 P35 Q43 R76 P77 G78 G79 E80 Q81 G82 K83 K91 D100 N101 Q111 M112 H113 D119 G125 K137 V145

• Molecule 2: Thanatin-like derivative

Chain B:

50%   12%   38%

V206 P207 LE1211 N212 R213 A214 T215 4F0216 K217 C218 A219 R220 Y221

4.2.14 Score per residue for model 14

• Molecule 1: Lipopolysaccharide export system protein LptA
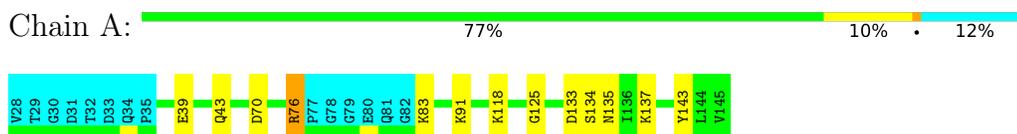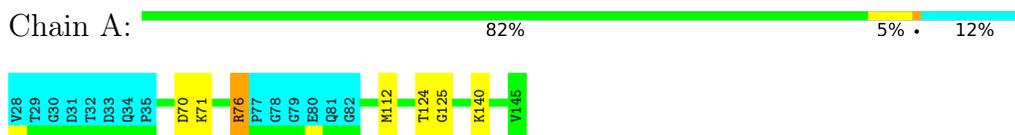
Chain A:

• Molecule 2: Thanatin-like derivative

Chain B:

4.2.15 Score per residue for model 15

• Molecule 1: Lipopolysaccharide export system protein LptA

Chain A:

• Molecule 2: Thanatin-like derivative

Chain B:

4.2.16 Score per residue for model 16

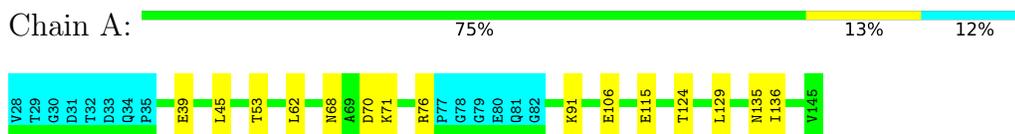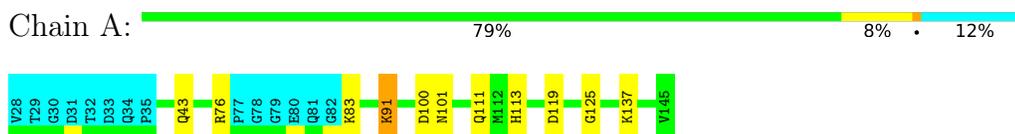• Molecule 1: Lipopolysaccharide export system protein LptA
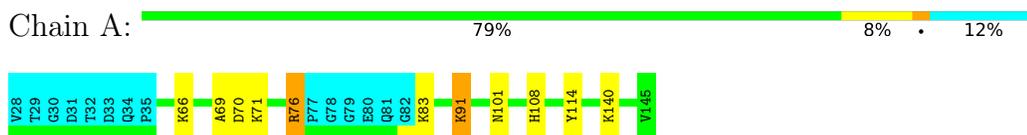
Chain A:

• Molecule 2: Thanatin-like derivative

Chain B:

4.2.17 Score per residue for model 17

• Molecule 1: Lipopolysaccharide export system protein LptA

Chain A:
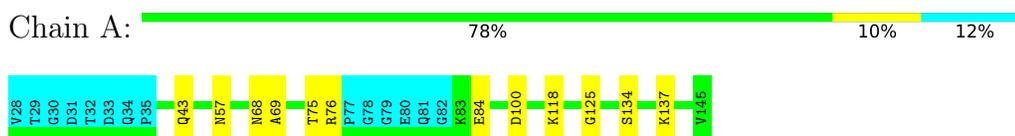

• Molecule 2: Thanatin-like derivative

Chain B:
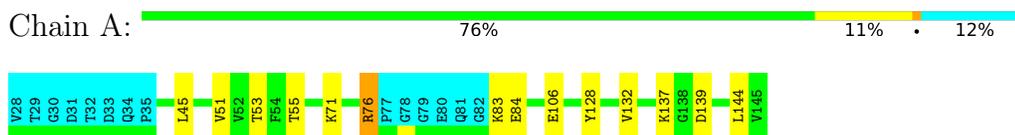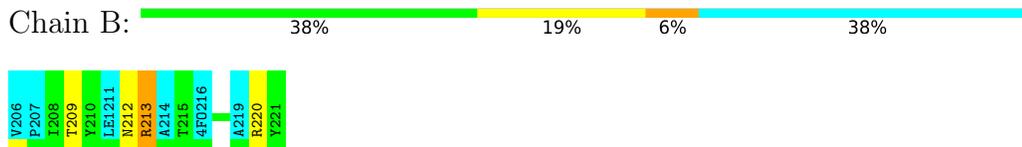
4.2.18 Score per residue for model 18

• Molecule 1: Lipopolysaccharide export system protein LptA

Chain A:

• Molecule 2: Thanatin-like derivative

Chain B:
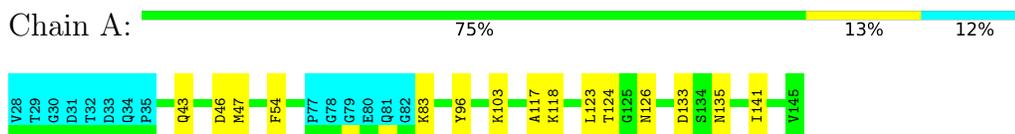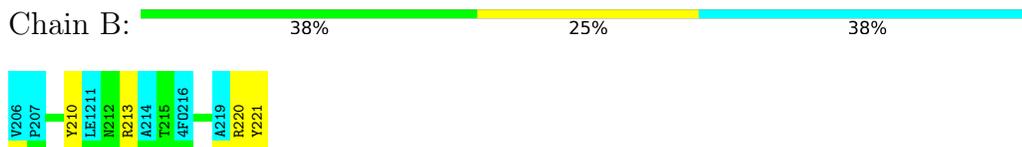
4.2.19 Score per residue for model 19

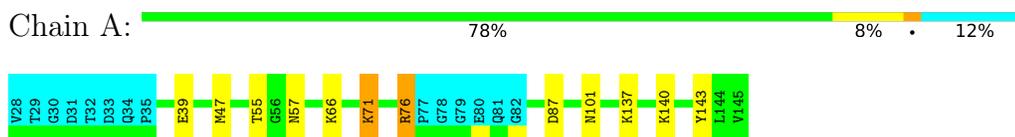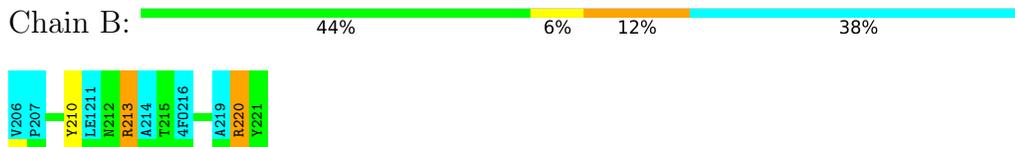• Molecule 1: Lipopolysaccharide export system protein LptA

Chain A:

• Molecule 2: Thanatin-like derivative

Chain B:
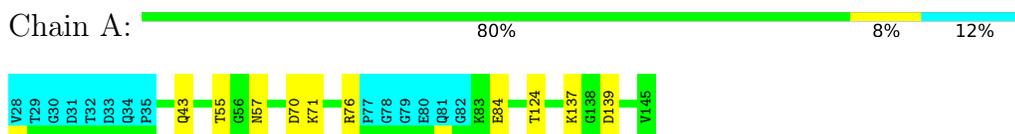
4.2.20 Score per residue for model 20

• Molecule 1: Lipopolysaccharide export system protein LptA

Chain A:

78%   9%   ·   12%

V28 T29 G30 D31 T32 D33 Q34 P35 I36 Q43 M47 T53 F54 T55 P77 G78 G79 E80 Q81 G82 K83 Y89 T94 V105 E106 E115 V132 V145

• Molecule 2: Thanatin-like derivative

Chain B:

44%   19%   38%

V206 P207 I208 T209 Y210 LE1211 A214 T215 4FO216 K217 C218 A219 R220 Y221
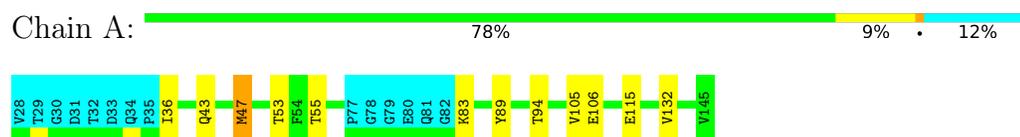
5   Refinement protocol and experimental data overview ⓘ

The models were refined using the following method: *na*.

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *structures with the lowest energy.*

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
MOE	refinement	2019.01
CYANA	structure calculation	3.98

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	1458
Number of shifts mapped to atoms	1458
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	83%

6 Model quality ⓘ

6.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: 4FO, DAB, HYP, EU0, LE1

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 (average) 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	0.66±0.01	0±0/820 (0.0± 0.0%)	1.00±0.03	1±1/1111 (0.1± 0.1%)
2	B	0.93±0.03	0±0/89 (0.0± 0.0%)	1.42±0.14	1±1/110 (1.1± 0.8%)
All	All	0.69	0/18180 (0.0%)	1.05	38/24420 (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	Planarity
1	A	0.0±0.0	0.8±0.9
2	B	0.0±0.0	0.5±0.6
All	All	0	27

There are no bond-length outliers.

All unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
2	B	220	ARG	NE-CZ-NH1	10.09	125.35	120.30	8	7
1	A	76	ARG	NE-CZ-NH1	8.73	124.67	120.30	8	10
2	B	213	ARG	NE-CZ-NH1	7.78	124.19	120.30	15	13
2	B	221	TYR	CB-CG-CD2	-6.77	116.94	121.00	1	1
2	B	213	ARG	CD-NE-CZ	5.94	131.91	123.60	17	1
1	A	76	ARG	NH1-CZ-NH2	-5.83	112.99	119.40	10	1
1	A	114	TYR	CB-CG-CD2	-5.71	117.58	121.00	9	2
2	B	213	ARG	NE-CZ-NH2	-5.59	117.51	120.30	19	1
1	A	76	ARG	NE-CZ-NH2	5.47	123.04	120.30	10	1
2	B	220	ARG	CD-NE-CZ	5.37	131.12	123.60	8	1

There are no chirality outliers.

All unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
2	B	220	ARG	Sidechain,Peptide	5
2	B	213	ARG	Sidechain	2
2	B	221	TYR	Sidechain	2
1	A	89	TYR	Sidechain	2
1	A	96	TYR	Sidechain	2
1	A	69	ALA	Peptide	2
1	A	99	GLN	Peptide	1
1	A	76	ARG	Peptide	1
1	A	118	LYS	Peptide	1
1	A	103	LYS	Peptide	1
1	A	139	ASP	Peptide	1
1	A	143	TYR	Sidechain	1
1	A	91	LYS	Peptide	1
1	A	57	ASN	Peptide	1
2	B	210	TYR	Sidechain	1
1	A	45	LEU	Peptide	1
1	A	46	ASP	Peptide	1
1	A	84	GLU	Peptide	1

6.2 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 each 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 averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
1	A	808	806	806	0±0
All	All	18000	17960	17960	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 0.

All unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Total
1:A:71:LYS:CE	1:A:71:LYS:HA	0.44	2.43	16	1

6.3 Torsion angles ⓘ

6.3.1 Protein backbone ⓘ

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all NMR entries. 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	103/118 (87%)	91±3 (89±3%)	10±3 (10±2%)	2±1 (2±1%)	13	56
2	B	9/16 (56%)	8±1 (91±9%)	1±1 (8±9%)	0±0 (1±3%)	18	66
All	All	2240/2680 (84%)	1990 (89%)	212 (9%)	38 (2%)	13	56

All 17 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	125	GLY	10
1	A	132	VAL	4
1	A	118	LYS	4
1	A	119	ASP	3
1	A	47	MET	3
1	A	117	ALA	2
1	A	91	LYS	2
1	A	50	ASN	1
1	A	100	ASP	1
1	A	92	PRO	1
2	B	220	ARG	1
1	A	84	GLU	1
1	A	101	ASN	1
1	A	134	SER	1
1	A	51	VAL	1
2	B	213	ARG	1
1	A	126	ASN	1

6.3.2 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 PDB entries followed by that with respect to all NMR entries. 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	90/100 (90%)	80±2 (89±2%)	10±2 (11±2%)	9	53
2	B	10/10 (100%)	9±1 (87±10%)	1±1 (13±10%)	7	48
All	All	2000/2200 (91%)	1772 (89%)	228 (11%)	9	52

All 66 unique residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	71	LYS	12
1	A	137	LYS	12
1	A	83	LYS	10
1	A	76	ARG	10
1	A	55	THR	9
1	A	70	ASP	9
1	A	124	THR	9
1	A	43	GLN	9
2	B	220	ARG	8
1	A	91	LYS	7
1	A	135	ASN	7
1	A	140	LYS	6
1	A	103	LYS	5
2	B	217	LYS	5
1	A	101	ASN	5
1	A	139	ASP	4
1	A	94	THR	4
1	A	106	GLU	4
1	A	118	LYS	4
2	B	213	ARG	4
1	A	53	THR	4
1	A	57	ASN	4
1	A	115	GLU	3
1	A	133	ASP	3
1	A	65	ILE	3
1	A	87	ASP	3
2	B	209	THR	3
1	A	100	ASP	3
1	A	66	LYS	3
1	A	128	TYR	3
2	B	212	ASN	3
1	A	68	ASN	3
1	A	84	GLU	3
1	A	39	GLU	3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	64	THR	2
1	A	99	GLN	2
2	B	221	TYR	2
1	A	41	ASP	2
1	A	62	LEU	2
1	A	75	THR	2
1	A	136	ILE	2
1	A	129	LEU	2
1	A	89	TYR	2
1	A	50	ASN	1
1	A	44	SER	1
1	A	48	GLN	1
1	A	130	GLN	1
1	A	131	GLN	1
1	A	86	ILE	1
1	A	98	MET	1
1	A	134	SER	1
1	A	112	MET	1
1	A	45	LEU	1
1	A	111	GLN	1
1	A	113	HIS	1
1	A	108	HIS	1
1	A	114	TYR	1
1	A	143	TYR	1
1	A	144	LEU	1
1	A	54	PHE	1
1	A	123	LEU	1
1	A	141	ILE	1
2	B	210	TYR	1
1	A	36	ILE	1
1	A	47	MET	1
1	A	105	VAL	1

6.3.3   RNA ⓘ

There are no RNA molecules in this entry.

6.4   Non-standard residues in protein, DNA, RNA chains ⓘ

6 non-standard protein/DNA/RNA residues are modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics

could be retrieved, the number of bonds that are observed in the model and the number of bonds 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 is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
2	DAB	B	214	2	5,6,7	0.60±0.10	0±0 (0±0%)
2	LE1	B	211	2	3,7,8	0.58±0.04	0±0 (0±0%)
2	HYP	B	207	2	6,8,9	0.50±0.04	0±0 (0±0%)
2	DAB	B	219	2	5,6,7	0.65±0.10	0±0 (0±0%)
2	4FO	B	216	2	5,6,7	0.59±0.09	0±0 (0±0%)
2	EU0	B	206	2	8,9,10	1.33±0.19	0±0 (5±6%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of 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 angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
2	DAB	B	214	2	1,6,8	0.27±0.16	0±0 (0±0%)
2	LE1	B	211	2	3,10,12	0.96±0.15	0±0 (0±0%)
2	HYP	B	207	2	5,10,12	1.04±0.20	0±0 (3±7%)
2	DAB	B	219	2	1,6,8	0.28±0.15	0±0 (0±0%)
2	4FO	B	216	2	1,6,8	0.33±0.21	0±0 (0±0%)
2	EU0	B	206	2	10,11,13	1.42±0.25	2±1 (18±8%)

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
2	LE1	B	211	2	-	0±0,4,8,10	-
2	4FO	B	216	2	-	0±0,4,5,7	-
2	EU0	B	206	2	-	0±0,9,10,12	-
2	DAB	B	219	2	-	0±0,4,5,7	-
2	HYP	B	207	2	-	0±0,0,11,13	0±0,1,1,1

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	DAB	B	214	2	-	0±0,4,5,7	-

All unique bond outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
2	B	206	EU0	CA-N	4.96	1.53	1.46	6	6
2	B	206	EU0	CT-N	2.08	1.29	1.33	8	2

All unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
2	B	206	EU0	CB-CA-N	4.77	117.39	111.17	20	4
2	B	206	EU0	CG1-CB-CA	3.08	115.92	111.21	19	1
2	B	206	EU0	CB-CA-C	3.01	116.82	113.04	20	2
2	B	206	EU0	CG2-CB-CA	2.88	115.62	111.21	5	2
2	B	206	EU0	CA-N-CT	2.76	131.21	123.19	19	5
2	B	207	HYP	CG-CB-CA	2.73	100.51	103.96	7	1
2	B	206	EU0	NT2-CT-N	2.44	124.99	119.55	6	6
2	B	207	HYP	O-C-CA	2.41	118.47	124.78	19	2
2	B	206	EU0	O-C-CA	2.37	118.21	124.83	12	15
2	B	206	EU0	N-CT-NT1	2.17	116.81	120.59	7	1

There are no chirality outliers.

All unique torsion outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Models (Total)
2	B	206	EU0	NT1-CT-N-CA	1
2	B	206	EU0	NT2-CT-N-CA	1

There are no ring outliers.

6.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.6 Ligand geometry ⓘ

There are no ligands in this entry.

6.7 Other polymers ⓘ

There are no such molecules in this entry.

6.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

7 Chemical shift validation ⓘ

The completeness of assignment taking into account all chemical shift lists is 83% for the well-defined parts and 83% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *starch_output*

7.1.1 Bookkeeping ⓘ

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	1458
Number of shifts mapped to atoms	1458
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

7.1.2 Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	118	-0.03 ± 0.11	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	105	-0.28 ± 0.17	None needed (< 0.5 ppm)
$^{13}C'$	105	0.45 ± 0.06	None needed (< 0.5 ppm)
^{15}N	110	-0.57 ± 0.21	Should be applied

7.1.3 Completeness of resonance assignments ⓘ

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 83%, i.e. 1295 atoms were assigned a chemical shift out of a possible 1556. 0 out of 19 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	535/575 (93%)	234/235 (100%)	200/228 (88%)	101/112 (90%)
Sidechain	704/867 (81%)	491/563 (87%)	207/271 (76%)	6/33 (18%)

Continued on next page...

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	56/114 (49%)	39/55 (71%)	17/56 (30%)	0/3 (0%)
Overall	1295/1556 (83%)	764/853 (90%)	424/555 (76%)	107/148 (72%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 83%, i.e. 1416 atoms were assigned a chemical shift out of a possible 1701. 0 out of 20 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	594/645 (92%)	261/265 (98%)	223/256 (87%)	110/124 (89%)
Sidechain	766/942 (81%)	532/610 (87%)	227/297 (76%)	7/35 (20%)
Aromatic	56/114 (49%)	39/55 (71%)	17/56 (30%)	0/3 (0%)
Overall	1416/1701 (83%)	832/930 (89%)	467/609 (77%)	117/162 (72%)

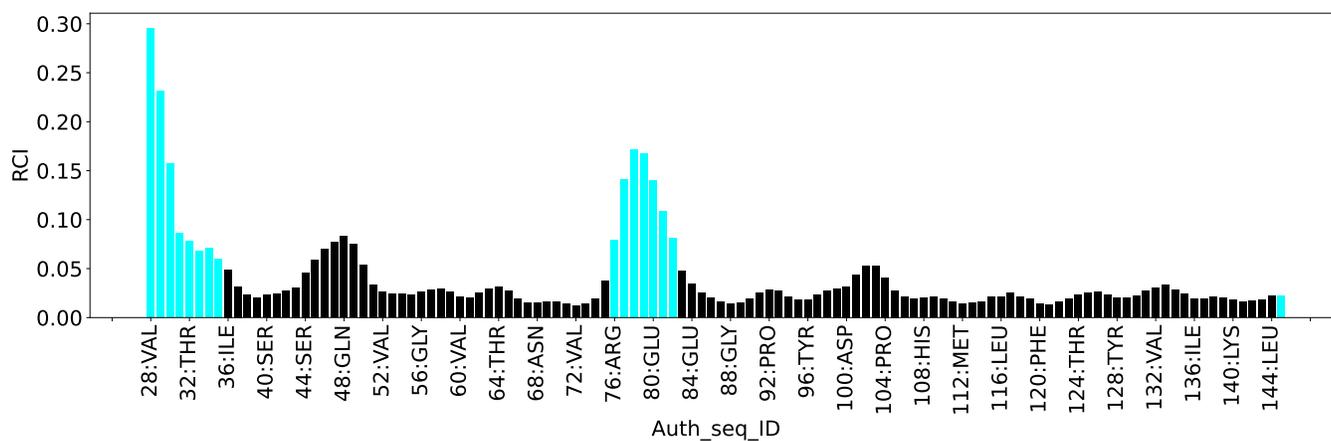### 7.1.4 Statistically unusual chemical shifts (i)

The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	113	HIS	CD2	136.74	103.95 – 136.66	5.0

7.1.5 Random Coil Index (RCI) plots (i)

The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A:

Random coil index (RCI) for chain B: