Full wwPDB NMR Structure Validation Report (i)

Mar 1, 2022 – 02:49 PM EST

PDB ID	:	2F8B
Title	:	NMR structure of the C-terminal domain (dimer) of HPV45 oncoprotein E7
Authors	:	Ohlenschlager, O.; Gorlach, M.
Deposited on	:	2005-12-02

This is a Full wwPDB NMR Structure Validation 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/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

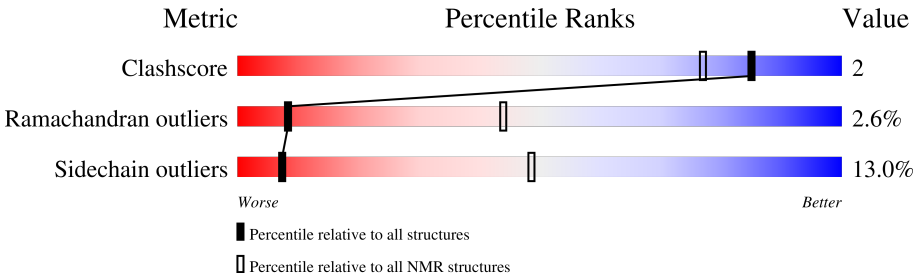
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.27

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 was not calculated.

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)	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	56	75% 9% • 14%
1	B	56	75% 12% • 11%

2 Ensemble composition and analysis

This entry contains 15 models. Model 1 is the overall representative, medoid model (most similar to other models).

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:8-A:55, B:7-B:56 (98)	0.37	1

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 4 clusters and 1 single-model cluster was found.

Cluster number	Models
1	1, 7, 12, 13, 15
2	4, 5, 8, 10
3	3, 6, 9
4	2, 11
Single-model clusters	14

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1724 atoms, of which 852 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Protein E7.

Mol	Chain	Residues	Atoms						Trace
1	A	56	Total	C	H	N	O	S	0
			861	269	426	77	82	7	
1	B	56	Total	C	H	N	O	S	0
			861	269	426	77	82	7	

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	cloning artifact	UNP P21736
A	2	SER	-	cloning artifact	UNP P21736
A	3	HIS	-	cloning artifact	UNP P21736
A	4	MET	-	cloning artifact	UNP P21736
B	1	GLY	-	cloning artifact	UNP P21736
B	2	SER	-	cloning artifact	UNP P21736
B	3	HIS	-	cloning artifact	UNP P21736
B	4	MET	-	cloning artifact	UNP P21736

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
2	A	1	Total	Zn
			1	1
2	B	1	Total	Zn
			1	1

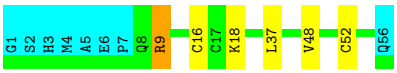
4 Residue-property plots

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.

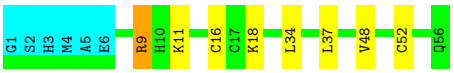
• Molecule 1: Protein E7

Chain A:



• Molecule 1: Protein E7

Chain B:



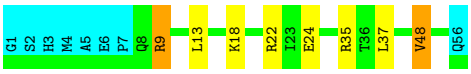
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1 (medoid)

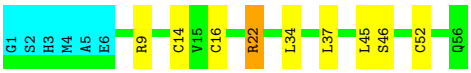
• Molecule 1: Protein E7

Chain A:



• Molecule 1: Protein E7

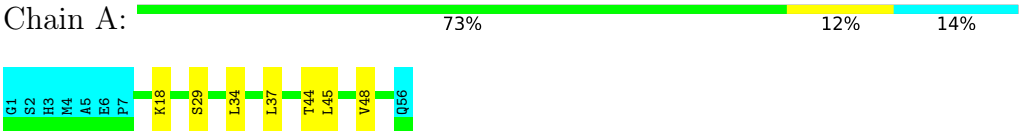
Chain B:



4.2.2 Score per residue for model 2

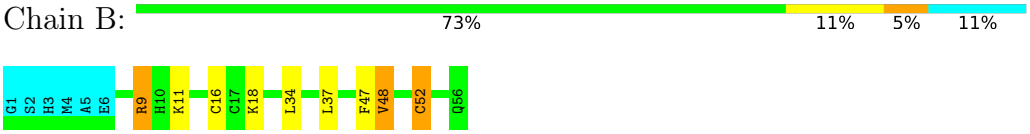
• Molecule 1: Protein E7

Chain A:



• Molecule 1: Protein E7

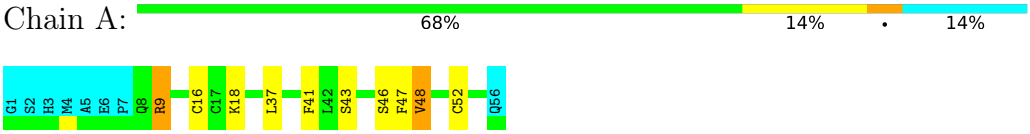
Chain B:



4.2.3 Score per residue for model 3

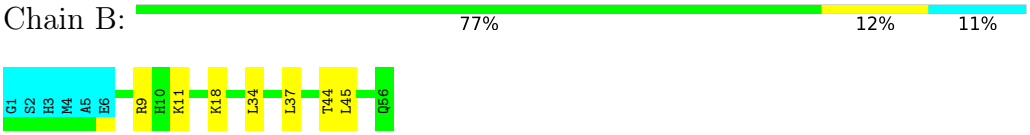
• Molecule 1: Protein E7

Chain A:



• Molecule 1: Protein E7

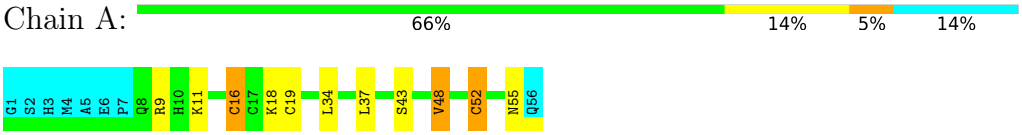
Chain B:



4.2.4 Score per residue for model 4

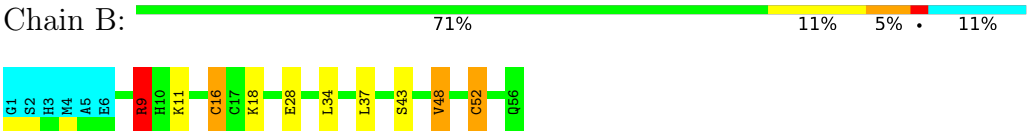
• Molecule 1: Protein E7

Chain A:



• Molecule 1: Protein E7

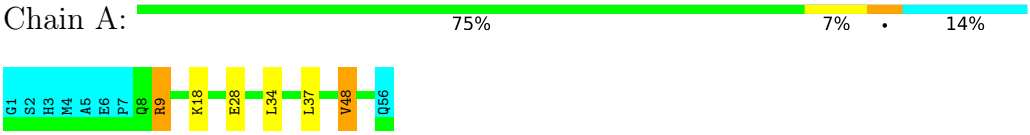
Chain B:



4.2.5 Score per residue for model 5

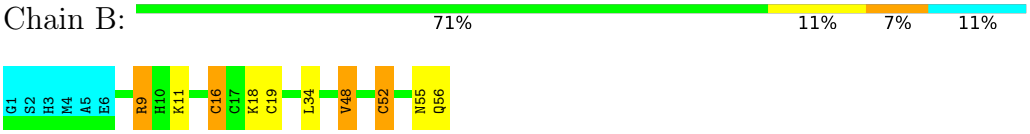
• Molecule 1: Protein E7

Chain A:



• Molecule 1: Protein E7

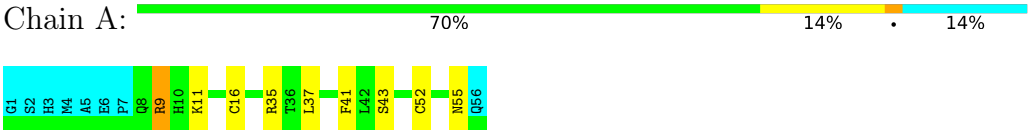
Chain B:



4.2.6 Score per residue for model 6

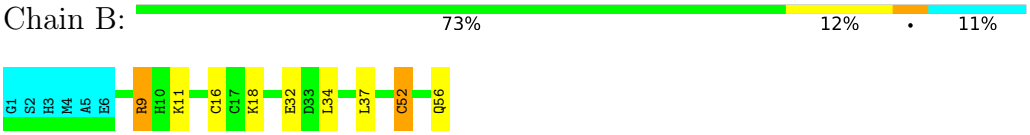
• Molecule 1: Protein E7

Chain A:



• Molecule 1: Protein E7

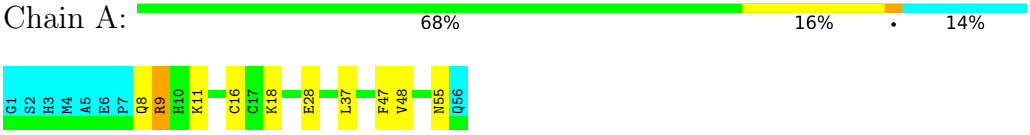
Chain B:



4.2.7 Score per residue for model 7

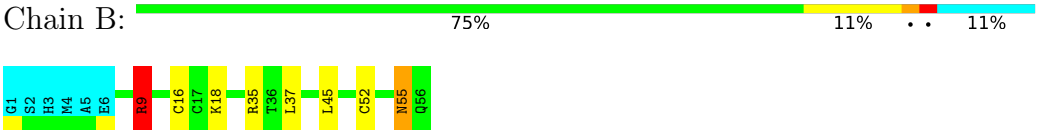
• Molecule 1: Protein E7

Chain A:



• Molecule 1: Protein E7

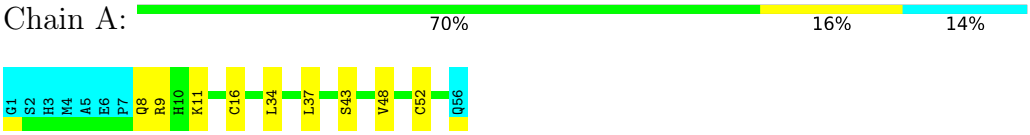
Chain B:



4.2.8 Score per residue for model 8

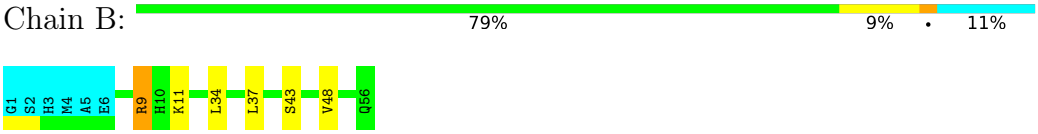
• Molecule 1: Protein E7

Chain A:



• Molecule 1: Protein E7

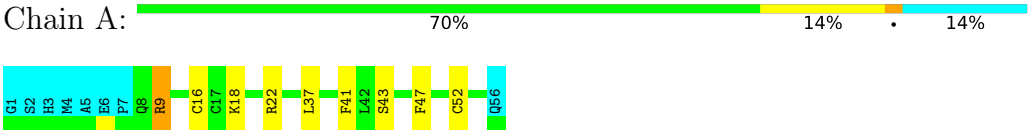
Chain B:



4.2.9 Score per residue for model 9

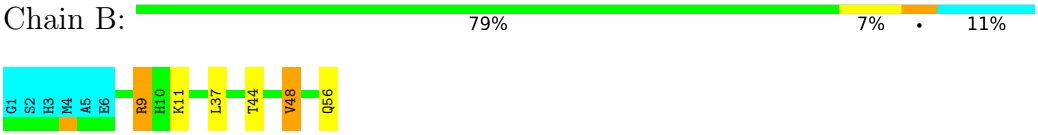
• Molecule 1: Protein E7

Chain A:



• Molecule 1: Protein E7

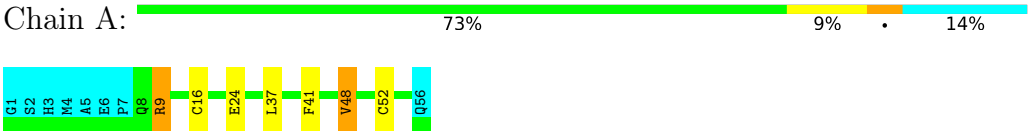
Chain B:



4.2.10 Score per residue for model 10

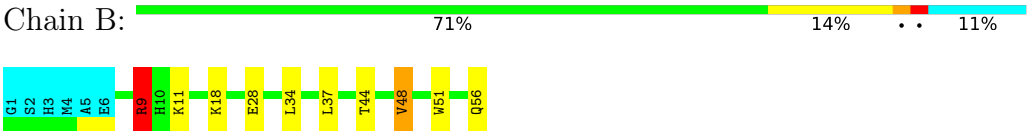
• Molecule 1: Protein E7

Chain A:



• Molecule 1: Protein E7

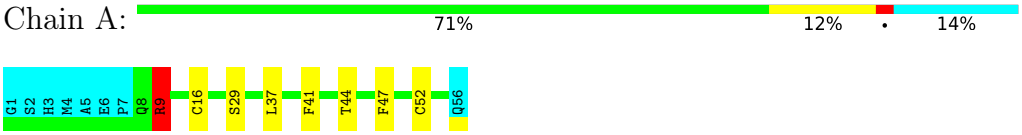
Chain B:



4.2.11 Score per residue for model 11

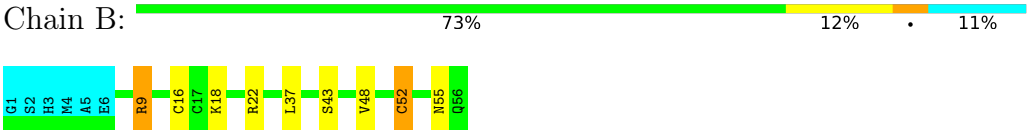
• Molecule 1: Protein E7

Chain A:



• Molecule 1: Protein E7

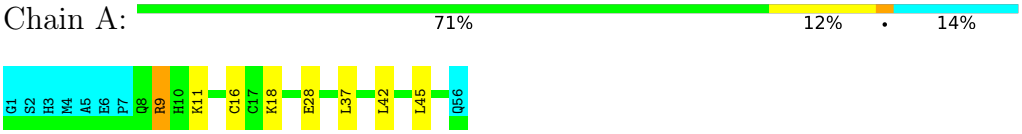
Chain B:



4.2.12 Score per residue for model 12

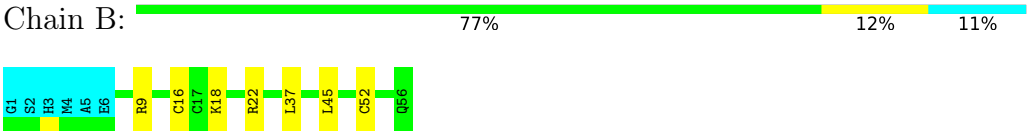
• Molecule 1: Protein E7

Chain A:



• Molecule 1: Protein E7

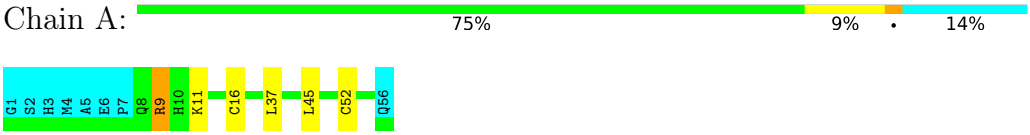
Chain B:



4.2.13 Score per residue for model 13

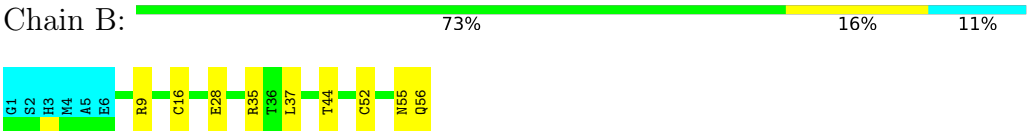
• Molecule 1: Protein E7

Chain A:



• Molecule 1: Protein E7

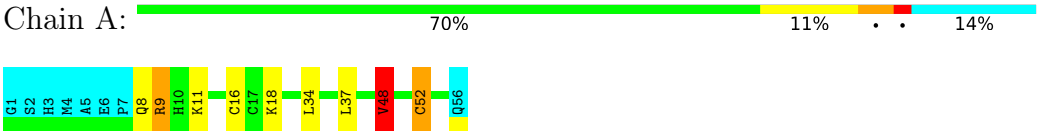
Chain B:



4.2.14 Score per residue for model 14

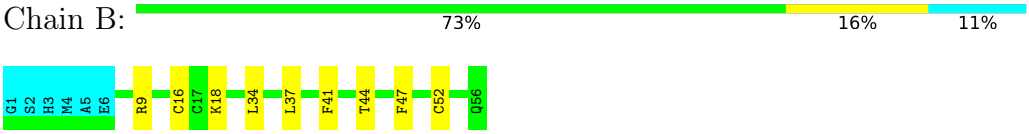
• Molecule 1: Protein E7

Chain A:



• Molecule 1: Protein E7

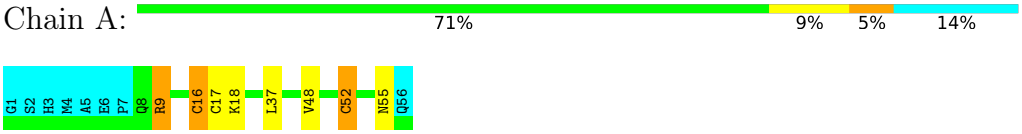
Chain B:



4.2.15 Score per residue for model 15

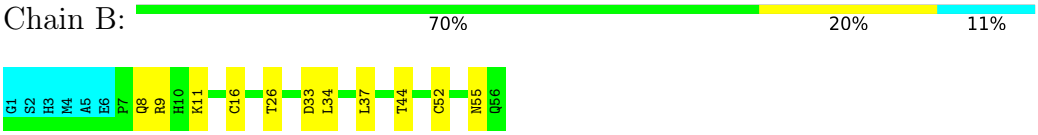
• Molecule 1: Protein E7

Chain A:



• Molecule 1: Protein E7

Chain B:



5 Refinement protocol and experimental data overview

The models were refined using the following method: *distance geometry/simulated annealing by CYANA, energy minimisation by OPAL.*

Of the 100 calculated structures, 15 were deposited, based on the following criterion: *target function.*

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

Software name	Classification	Version
CYANA	structure solution	2.1
OPAL	refinement	2.6

No chemical shift data was provided.

6 Model quality

6.1 Standard geometry

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 (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	0.57±0.02	0±0/383 (0.0± 0.0%)	1.15±0.06	1±1/519 (0.3± 0.2%)
1	B	0.56±0.01	0±0/399 (0.0± 0.0%)	1.14±0.06	1±1/539 (0.2± 0.2%)
All	All	0.56	0/11730 (0.0%)	1.14	36/15870 (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	B	0.0±0.0	0.1±0.3
All	All	0	2

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
1	B	48	VAL	CA-CB-CG1	9.53	125.20	110.90	5	3
1	A	48	VAL	CA-CB-CG1	8.81	124.12	110.90	4	5
1	B	48	VAL	CG1-CB-CG2	-8.67	97.03	110.90	5	4
1	A	48	VAL	CG1-CB-CG2	-8.17	97.83	110.90	4	1
1	A	9	ARG	NE-CZ-NH1	7.29	123.94	120.30	15	2
1	A	16	CYS	CA-CB-SG	-6.25	102.74	114.00	4	2
1	B	16	CYS	CA-CB-SG	-6.17	102.90	114.00	5	2
1	A	28	GLU	CB-CA-C	5.95	122.31	110.40	7	2
1	A	48	VAL	CB-CA-C	5.64	122.13	111.40	5	3
1	B	9	ARG	NE-CZ-NH2	-5.60	117.50	120.30	7	2
1	B	35	ARG	NE-CZ-NH1	5.28	122.94	120.30	13	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	B	48	VAL	CB-CA-C	5.28	121.42	111.40	10	1
1	A	35	ARG	NE-CZ-NH1	5.27	122.94	120.30	6	2
1	A	9	ARG	NE-CZ-NH2	-5.13	117.73	120.30	10	1
1	A	55	ASN	C-N-CA	5.04	134.29	121.70	6	1
1	A	48	VAL	CA-CB-CG2	5.02	118.43	110.90	14	1
1	B	28	GLU	CB-CA-C	5.01	120.42	110.40	13	1
1	B	55	ASN	C-N-CA	5.01	134.22	121.70	7	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)
1	B	22	ARG	Sidechain	1
1	B	9	ARG	Sidechain	1

6.2 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 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	B	393	391	392	2±2
1	A	377	376	377	2±1
2	B	1	0	0	1±0
2	A	1	0	0	1±0
All	All	11580	11505	11535	52

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 2.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:B:16:CYS:HG	2:B:57:ZN:ZN	0.89	0.61	5	10
1:A:16:CYS:HG	2:A:57:ZN:ZN	0.88	0.61	14	9
1:A:48:VAL:HG23	1:B:26:THR:HG22	0.60	1.72	15	1
1:A:48:VAL:CG2	1:B:26:THR:HG22	0.56	2.30	15	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:48:VAL:HB	1:A:52:CYS:SG	0.52	2.45	15	2
1:B:16:CYS:SG	1:B:52:CYS:SG	0.51	3.08	4	11
1:A:16:CYS:SG	1:A:52:CYS:SG	0.49	3.10	4	9
1:A:28:GLU:HG3	1:B:48:VAL:CG2	0.47	2.38	5	1
1:A:48:VAL:HG21	1:B:28:GLU:CD	0.47	2.30	10	1
1:A:48:VAL:CG1	1:A:52:CYS:SG	0.46	3.04	4	1
1:B:48:VAL:CG1	1:B:52:CYS:SG	0.45	3.04	4	2
1:A:28:GLU:CG	1:B:48:VAL:HG21	0.44	2.42	5	1
1:A:48:VAL:HG13	1:A:52:CYS:SG	0.44	2.52	3	1
1:A:28:GLU:HG3	1:B:48:VAL:HG21	0.43	1.91	5	1
1:A:48:VAL:CG2	1:B:28:GLU:HG3	0.41	2.45	4	1

6.3 Torsion angles (i)

6.3.1 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 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	48/56 (86%)	41±2 (86±4%)	6±2 (12±3%)	1±1 (3±2%)	8	44
1	B	49/56 (88%)	41±1 (84±1%)	7±1 (13±2%)	1±0 (3±1%)	8	44
All	All	1455/1680 (87%)	1234 (85%)	183 (13%)	38 (3%)	8	44

All 6 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	B	9	ARG	15
1	A	9	ARG	12
1	A	43	SER	5
1	B	43	SER	3
1	A	8	GLN	2
1	B	44	THR	1

6.3.2 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 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	45/51 (88%)	39±1 (87±3%)	6±1 (13±3%)	8	50
1	B	47/51 (92%)	41±1 (87±3%)	6±1 (13±3%)	7	48
All	All	1380/1530 (90%)	1201 (87%)	179 (13%)	7	48

All 42 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	37	LEU	15
1	B	37	LEU	14
1	A	9	ARG	12
1	A	18	LYS	10
1	B	34	LEU	10
1	B	18	LYS	10
1	B	9	ARG	9
1	B	11	LYS	9
1	A	11	LYS	7
1	A	34	LEU	5
1	B	48	VAL	5
1	B	52	CYS	5
1	A	41	PHE	5
1	B	44	THR	5
1	B	55	ASN	5
1	B	56	GLN	5
1	A	48	VAL	4
1	B	45	LEU	4
1	A	47	PHE	4
1	B	22	ARG	3
1	A	45	LEU	3
1	A	52	CYS	3
1	A	55	ASN	3
1	A	22	ARG	2
1	A	24	GLU	2
1	A	29	SER	2
1	A	44	THR	2
1	B	47	PHE	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	13	LEU	1
1	B	14	CYS	1
1	B	46	SER	1
1	A	46	SER	1
1	A	19	CYS	1
1	B	19	CYS	1
1	B	32	GLU	1
1	A	8	GLN	1
1	B	51	TRP	1
1	A	42	LEU	1
1	B	41	PHE	1
1	A	17	CYS	1
1	B	8	GLN	1
1	B	33	ASP	1

6.3.3 RNA

There are no RNA molecules in this entry.

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

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

6.5 Carbohydrates

There are no monosaccharides in this entry.

6.6 Ligand geometry

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

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

No chemical shift data were provided