Full wwPDB NMR Structure Validation Report (i)

May 28, 2020 – 07:24 pm BST

PDB ID	:	1E7J
Title	:	HMG-D complexed to a bulge DNA
Authors	:	Cerdan, R.; Payet, D.; Yang, J.-C.; Travers, A.A.; Neuhaus, D.
Deposited on	:	2000-08-29

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 (1)) were used in the production of this report:

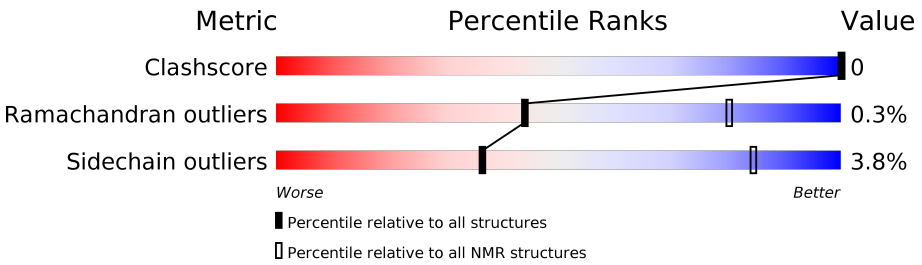
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

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	74	86% 14%
2	B	14	100%
3	C	12	92% 8%

2 Ensemble composition and analysis (i)

This entry contains 10 models. Model 8 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative.

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:71 (64)	0.82	8

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 4 single-model clusters were found.

Cluster number	Models
1	2, 5, 8, 10
2	1, 3
Single-model clusters	4; 6; 7; 9

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 2021 atoms, of which 897 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called HIGH MOBILITY GROUP PROTEIN D.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	74	1197	372	601	109	112	3	0

- Molecule 2 is a DNA chain called DNA (5'-D(*CP*GP*AP*TP*AP*TP*TP*AP*AP*GP*AP*GP*CP*C)-3').

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	14	444	137	159	55	80	13	0

- Molecule 3 is a DNA chain called DNA (5'-D(*GP*GP*CP*TP*CP*AP*AP*TP*AP*TP*CP*G)-3').

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
3	C	12	380	117	137	45	70	11	0

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA 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: HIGH MOBILITY GROUP PROTEIN D

Chain A: 86% 14%

M1 S2 D3 K4 P5 K6 R7 A72 N73 G74

• Molecule 2: DNA (5'-D(*CP*GP*AP*TP*AP*TP*TP*AP*AP*GP*AP*GP*CP*C)-3')

Chain B: 100%

C1 G2 A3 T4 A5 T6 T7 A8 A9 G10 A11 G12 C13 C14

• Molecule 3: DNA (5'-D(*GP*GP*CP*TP*CP*AP*AP*TP*AP*TP*CP*G)-3')

Chain C: 92% 8%

G15 G16 C17 T18 C19 A20 A21 T22 A23 T24 C25 G26

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: HIGH MOBILITY GROUP PROTEIN D

Chain A: 84% • 14%

M1 S2 D3 K4 P5 K6 R7 K24 M46 A72 N73 G74

• Molecule 2: DNA (5'-D(*CP*GP*AP*TP*AP*TP*TP*AP*AP*GP*AP*GP*CP*C)-3')

Chain B: 7% 71% 21%

C1 G2 A3 T4 A5 T6 T7 A8 A9 G10 A11 G12 C13 C14

- Molecule 3: DNA (5'-D(*GP*GP*CP*TP*CP*AP*AP*TP*AP*TP*CP*G)-3')

Chain C: 100%

G15 G16 C17 T18 C19 A20 A21 T22 A23 T24 C25 G26

4.2.2 Score per residue for model 2

- Molecule 1: HIGH MOBILITY GROUP PROTEIN D

Chain A: 84% • 14%

M1 S2 D3 K4 P5 K6 R7 M46 K49 A72 N73 G74

- Molecule 2: DNA (5'-D(*CP*GP*AP*TP*AP*TP*TP*AP*AP*GP*AP*GP*CP*C)-3')

Chain B: 93% 7%

C1 G2 A3 T4 A5 T6 T7 A8 A9 G10 A11 G12 C13 C14

- Molecule 3: DNA (5'-D(*GP*GP*CP*TP*CP*AP*AP*TP*AP*TP*CP*G)-3')

Chain C: 8% 83% 8%

G15 G16 C17 T18 C19 A20 A21 T22 A23 T24 C25 G26

4.2.3 Score per residue for model 3

- Molecule 1: HIGH MOBILITY GROUP PROTEIN D

Chain A: 86% 14%

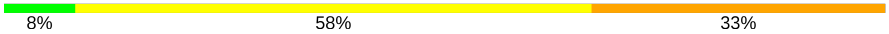
M1 S2 D3 K4 P5 K6 R7 A72 N73 G74

- Molecule 2: DNA (5'-D(*CP*GP*AP*TP*AP*TP*TP*AP*AP*GP*AP*GP*CP*C)-3')

Chain B: 7% 79% 14%

• Molecule 3: DNA (5'-D(*GP*GP*CP*TP*CP*AP*AP*TP*AP*TP*CP*G)-3')

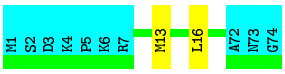
Chain C:



4.2.4 Score per residue for model 4

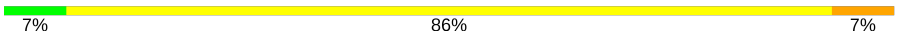
• Molecule 1: HIGH MOBILITY GROUP PROTEIN D

Chain A:



• Molecule 2: DNA (5'-D(*CP*GP*AP*TP*AP*TP*TP*AP*AP*GP*AP*GP*CP*C)-3')

Chain B:



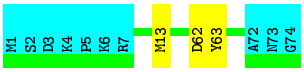
• Molecule 3: DNA (5'-D(*GP*GP*CP*TP*CP*AP*AP*TP*AP*TP*CP*G)-3')

Chain C:

4.2.5 Score per residue for model 5

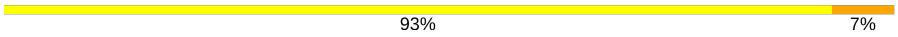
• Molecule 1: HIGH MOBILITY GROUP PROTEIN D

Chain A:



• Molecule 2: DNA (5'-D(*CP*GP*AP*TP*AP*TP*TP*AP*AP*GP*AP*GP*CP*C)-3')

Chain B:



• Molecule 3: DNA (5'-D(*GP*GP*CP*TP*CP*AP*AP*TP*AP*TP*CP*G)-3')

Chain C: 8% 75% 17%

G15 G16 C17 T18 C19 A20 A21 T22 A23 T24 C25 G26

4.2.6 Score per residue for model 6

- Molecule 1: HIGH MOBILITY GROUP PROTEIN D

Chain A: 85% • 14%

M1 S2 D3 K4 P5 K6 R7 P8 L9 A72 N73 G74

- Molecule 2: DNA (5'-D(*CP*GP*AP*TP*AP*TP*TP*AP*AP*GP*AP*GP*CP*C)-3')

Chain B: 100%

C1 G2 A3 T4 A5 T6 T7 A8 A9 G10 A11 G12 C13 C14

- Molecule 3: DNA (5'-D(*GP*GP*CP*TP*CP*AP*AP*TP*AP*TP*CP*G)-3')

Chain C: 8% 83% 8%

G15 G16 C17 T18 C19 A20 A21 T22 A23 T24 C25 G26

4.2.7 Score per residue for model 7

- Molecule 1: HIGH MOBILITY GROUP PROTEIN D

Chain A: 82% • 14%

M1 S2 D3 K4 P5 K6 R7 P8 L9 R44 K49 A72 N73 G74

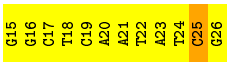
- Molecule 2: DNA (5'-D(*CP*GP*AP*TP*AP*TP*TP*AP*AP*GP*AP*GP*CP*C)-3')

Chain B: 7% 86% 7%

C1 G2 A3 T4 A5 T6 T7 A8 A9 G10 A11 G12 C13 C14

- Molecule 3: DNA (5'-D(*GP*GP*CP*TP*CP*AP*AP*TP*AP*TP*CP*G)-3')

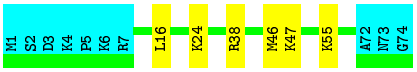
Chain C: 92% 8%



4.2.8 Score per residue for model 8 (medoid)

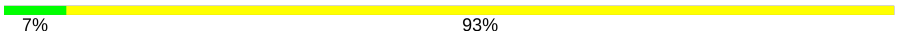
- Molecule 1: HIGH MOBILITY GROUP PROTEIN D

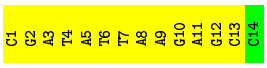
Chain A:



- Molecule 2: DNA (5'-D(*CP*GP*AP*TP*AP*TP*TP*AP*AP*GP*AP*GP*CP*C)-3')

Chain B:





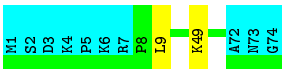
- Molecule 3: DNA (5'-D(*GP*GP*CP*TP*CP*AP*AP*TP*AP*TP*CP*G)-3')

Chain C:

4.2.9 Score per residue for model 9

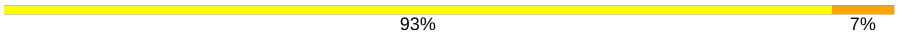
- Molecule 1: HIGH MOBILITY GROUP PROTEIN D

Chain A:



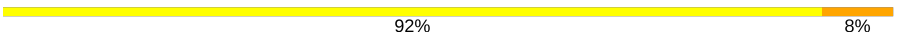
- Molecule 2: DNA (5'-D(*CP*GP*AP*TP*AP*TP*TP*AP*AP*GP*AP*GP*CP*C)-3')

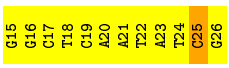
Chain B:



- Molecule 3: DNA (5'-D(*GP*GP*CP*TP*CP*AP*AP*TP*AP*TP*CP*G)-3')

Chain C:

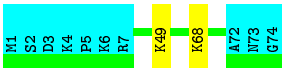




4.2.10 Score per residue for model 10

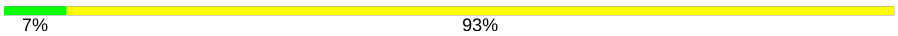
- Molecule 1: HIGH MOBILITY GROUP PROTEIN D

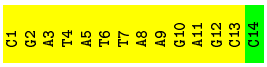
Chain A:



- Molecule 2: DNA (5'-D(*CP*GP*AP*TP*AP*TP*TP*AP*AP*GP*AP*GP*CP*C)-3')

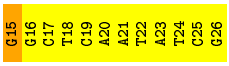
Chain B:





- Molecule 3: DNA (5'-D(*GP*GP*CP*TP*CP*AP*AP*TP*AP*TP*CP*G)-3')

Chain C:



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *SIMULATED ANNEALING, MINIMIZATION.*

Of the 50 calculated structures, 10 were deposited, based on the following criterion: *LOW VALUES OF THE NOE ENERGY TERM AND PROTEIN LOCATED WITHIN MINOR GROOVE OF DNA.*

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

Software name	Classification	Version
X-PLOR 3.851, CHARMM	refinement	24
XPLOR	structure solution	3.851
CHARMM	structure solution	24

No chemical shift data was provided. No validations of the models with respect to experimental NMR restraints is performed at this time.

COVALENT-GEOMETRY INFOmissingINFO

5.1 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
All	All	10480	8170	8170	-

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 -.

There are no clashes.

5.2 Torsion angles (i)

5.2.1 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	64/74 (86%)	60±1 (94±2%)	4±2 (6±3%)	0±0 (0±1%)	44 80
All	All	640/740 (86%)	601 (94%)	37 (6%)	2 (0%)	44 80

All 2 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	47	LYS	1
1	A	9	LEU	1

5.2.2 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	52/60 (87%)	50±1 (96±3%)	2±1 (4±3%)	36 84
All	All	520/600 (87%)	500 (96%)	20 (4%)	36 84

All 11 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	49	LYS	4
1	A	46	MET	3
1	A	24	LYS	2
1	A	16	LEU	2
1	A	13	MET	2
1	A	9	LEU	2
1	A	62	ASP	1
1	A	55	LYS	1
1	A	44	ARG	1
1	A	38	ARG	1
1	A	68	LYS	1

5.2.3 RNA (i)

There are no RNA molecules in this entry.

5.3 Non-standard residues in protein, DNA, RNA chains (i)

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

5.4 Carbohydrates (i)

There are no carbohydrates in this entry.

5.5 Ligand geometry (i)

There are no ligands in this entry.

5.6 Other polymers (i)

There are no such molecules in this entry.

5.7 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Chemical shift validation (i)

No chemical shift data were provided