wwPDB NMR Structure Validation Summary Report (i)

Feb 13, 2022 – 11:05 PM EST

PDB ID	:	1HRZ
Title	:	THE 3D STRUCTURE OF THE HUMAN SRY-DNA COMPLEX SOLVED BY MULTI-DIMENSIONAL HETERONUCLEAR-EDITED AND-FILTERED NMR
Authors	:	Clore, G.M.; Werner, M.H.; Huth, J.R.; Gronenborn, A.M.
Deposited on	:	1995-05-09

This is a wwPDB NMR 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/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:

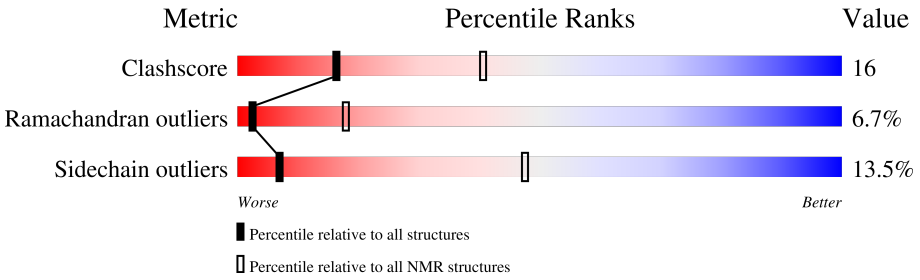
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.26
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.26

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	B	8	38% 62%
2	C	8	12% 50% 38%
3	A	76	50% 38% 5% . .

2 Ensemble composition and analysis (i)

This entry contains 35 models. Model 27 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:5-A:75 (71)	0.42	27

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 9 clusters and 3 single-model clusters were found.

Cluster number	Models
1	18, 22, 23, 27, 28, 29
2	8, 16, 17, 25, 26, 33
3	3, 11, 13, 15, 34
4	6, 14, 30, 35
5	1, 5, 31
6	9, 20
7	12, 24
8	7, 32
9	2, 10
Single-model clusters	4; 19; 21

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 1794 atoms, of which 828 are hydrogens and 0 are deuteriums.

- Molecule 1 is a DNA chain called DNA (5'-D(*GP*CP*AP*CP*AP*AP*AP*C)-3').

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	B	8	250	77	90	34	42	7	0

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	C	8	256	79	94	26	50	7	0

- Molecule 3 is a protein called HUMAN SRY.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
3	A	73	1288	409	644	123	107	5	0

4 Residue-property plots (i)

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

Chain B: 38% 62%

G1 C2 A3 C4 A5 A6 A7 C8

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

Chain C: 12% 50% 38%

G9 T10 T11 T12 G13 T14 G15 C16

• Molecule 3: HUMAN SRY

Chain A: 50% 38% 5%

VAL GLN D3 R4 V5 K6 R7 N10 A11 F12 W15 S16 R17 D18 Q19 R20 N27 P28 R29 M30 R31 N32 S33 E34 I35 S36 K37 Q42 W43 K44 M45 E50 K51 W52 P53 F54 F55 Q56 K60 R66 Y69 P70 R75 PRO

4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 27. Colouring as in section 4.1 above.

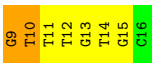
• Molecule 1: DNA (5'-D(*GP*CP*AP*CP*AP*AP*AP*C)-3')

Chain B: 38% 62%

G1 C2 A3 C4 A5 A6 A7 C8

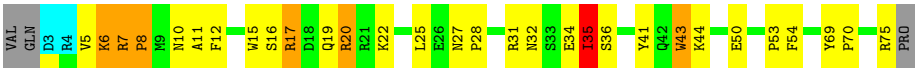
• Molecule 2: DNA (5'-D(*GP*TP*TP*TP*GP*TP*GP*C)-3')

Chain C: 12% 62% 25%



- Molecule 3: HUMAN SRY

Chain A:



5 Refinement protocol and experimental data overview (i)

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

Of the ? calculated structures, 35 were deposited, based on the following criterion: *?*.

The authors did not provide any information on software used for structure solution, optimization or refinement.

No chemical shift data was provided.

6 Model quality (i)

6.1 Standard geometry (i)

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	B	1.39±0.02	0±0/180 (0.0± 0.1%)	2.88±0.01	16±1/275 (6.0± 0.2%)
2	C	1.55±0.02	3±1/180 (1.8± 0.5%)	2.67±0.01	15±1/277 (5.5± 0.5%)
3	A	1.11±0.01	0±0/642 (0.0± 0.0%)	0.96±0.01	0±0/858 (0.0± 0.0%)
All	All	1.25	118/35070 (0.3%)	1.89	1109/49350 (2.2%)

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	1.0±0.0	0.0±0.0
2	C	1.0±0.0	0.0±0.0
3	A	0.0±0.0	0.0±0.2
All	All	70	1

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	C	14	DT	C5-C7	6.52	1.53	1.50	11	35
2	C	11	DT	C5-C7	6.46	1.53	1.50	27	32
2	C	10	DT	C5-C7	6.07	1.53	1.50	8	26
2	C	12	DT	C5-C7	6.00	1.53	1.50	18	23
1	B	1	DG	N9-C8	-5.21	1.34	1.37	8	2

5 of 43 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	1	DG	N7-C8-N9	14.37	120.28	113.10	25	35

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
2	C	13	DG	N7-C8-N9	14.29	120.24	113.10	3	35
2	C	15	DG	N7-C8-N9	14.17	120.18	113.10	33	35
2	C	9	DG	N7-C8-N9	14.14	120.17	113.10	9	35
1	B	6	DA	N7-C8-N9	12.82	120.21	113.80	7	35

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

Mol	Chain	Res	Type	Atoms	Models (Total)
1	B	5	DA	C3'	35
2	C	9	DG	C3'	35

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
3	A	7	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	160	90	90	1±1
2	C	162	94	94	2±1
3	A	625	627	627	26±5
All	All	33145	28385	28385	975

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 16.

5 of 236 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
3:A:19:GLN:NE2	3:A:39:LEU:HD21	0.74	1.97	28	2
3:A:7:ARG:NE	3:A:7:ARG:H	0.70	1.84	14	8
3:A:5:VAL:HG12	3:A:6:LYS:H	0.68	1.46	7	7
3:A:56:GLN:NE2	3:A:56:GLN:N	0.67	2.43	16	9

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
3:A:56:GLN:H	3:A:56:GLN:NE2	0.66	1.89	28	3

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	
3	A	70/76 (92%)	54±2 (78±3%)	11±2 (16±3%)	5±1 (7±2%)	2	18
All	All	2450/2660 (92%)	1904 (78%)	382 (16%)	164 (7%)	2	18

5 of 16 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
3	A	53	PRO	34
3	A	17	ARG	25
3	A	70	PRO	25
3	A	35	ILE	16
3	A	32	ASN	13

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	
3	A	65/70 (93%)	56±2 (86±3%)	9±2 (14±3%)	7	47
All	All	2275/2450 (93%)	1967 (86%)	308 (14%)	7	47

5 of 30 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)
3	A	7	ARG	35
3	A	35	ILE	35
3	A	43	TRP	35
3	A	31	ARG	21
3	A	6	LYS	20

6.3.3 RNA (i)

There are no RNA molecules in this entry.

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

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

6.5 Carbohydrates (i)

There are no monosaccharides in this entry.

6.6 Ligand geometry (i)

There are no ligands in this entry.

6.7 Other polymers (i)

There are no such molecules in this entry.

6.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

7 Chemical shift validation (i)

No chemical shift data were provided