



wwPDB NMR Structure Validation Summary Report

Oct 17, 2021 – 06:18 AM EDT

PDB ID : 1LFU
Title : NMR Solution Structure of the Extended PBX Homeodomain Bound to DNA
Authors : Sprules, T.; Green, N.; Featherstone, M.; Gehring, K.
Deposited on : 2002-04-12

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  symbol.

The following versions of software and data (see [references](#) ) 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 : 2.23.2
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.23.2

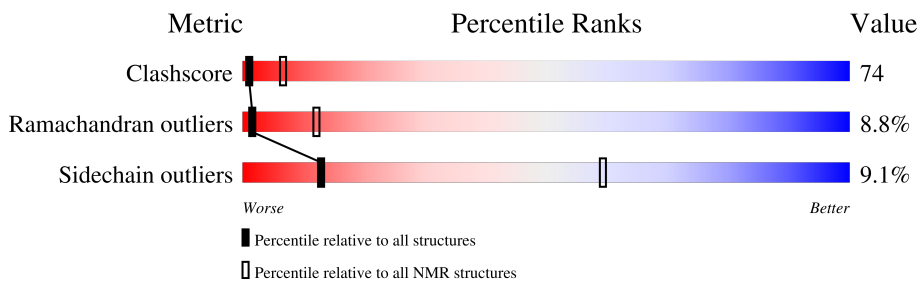
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	14	100%
2	B	14	100%
3	P	82	21% 52% 10% 17%

2 Ensemble composition and analysis i

This entry contains 20 models. Model 10 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

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	P:6-P:70 (65)	0.22	10

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 3 clusters. No single-model clusters were found.

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

3 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 2239 atoms, of which 995 are hydrogens and 0 are deuteriums.

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	A	14	441	135	159	51	83	13	0

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	14	444	136	158	56	81	13	0

- Molecule 3 is a protein called homeobox protein PBX1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
3	P	82	1354	426	678	124	125	1	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
P	0	MET	-	initiating methionine	UNP P41778
P	38	SER	CYS	engineered mutation	UNP P41778

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

Chain A: 

G1
C2
G3
C4
A5
T6
G7
A8
T9
T10
G11
C12
C13
C14

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

Chain B: 

G15
G16
G17
G18
A19
A20
T21
C22
A23
T24
G25
C26
G27
C28

- Molecule 3: homeobox protein PBX1

Chain P: 

M0
A1
R2
R3
K4
R5
R6
N7
F8
N9
K10
Q11
A12
T13
E14
I15
L16
M17
E18
Y19
F20
Y21
S22
H23
L23A
S23B
N23C
P24
Y25
P26
S27
E28
E29
A30
K31
E32
E33
L34
A35
K36
K37
S38
G39
I40
T41
V42
S43
Q44
V45
S46
M47
W48
F49
K52
R53
I54
R55
Y56
K57

K58
M59
I60
G61
K62
F63
E66
A67
I69
Y70
A71
A72
K73
T74
V75
V76
T77
A78

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

The representative model is number 10. Colouring as in section 4.1 above.

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

Chain A: 


G1
C2
G3
C4
A5
T6
G7
A8
T9
T10
G11
C12
C13
C14

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

Chain B:  100%

G15
G16
G17
C18
A19
A20
T21
C22
A23
T24
G25
C26
C27
C28

- Molecule 3: homeobox protein PBX1

Chain P:  23% 46% 13% 17%

N0
A1
R2
R3
K4
R5
R6
N7
F8
N9
K10
G11
A12
T13
E14
I15
L16
N17
E18
Y19
F20
Y21
S22
H23
L23A
S23B
N23C
P24
Y25
P26
S27
E28
E29
A30
K31
E32
E33
L34
A35
K36
K37
S38
G39
I40
T41
V42
S43
Q44
V45
W46
F49
K52
R53
I54
R55
Y56
K57
K58

N59
I60
G61
K62
F63
A67
Y70
A71
K72
T73
T74
A75
V76
T77
A78

5 Refinement protocol and experimental data overview

The models were refined using the following method: *dynamical annealing*.

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

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

Software name	Classification	Version
CNS	refinement	0.9, 1.1

No chemical shift data was provided.

6 Model quality [i](#)

6.1 Standard geometry [i](#)

There are no covalent bond-length or bond-angle outliers.

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
2	B	0.0±0.0	0.1±0.2
All	All	0	1

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
2	B	24	DT	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	A	282	159	159	38±4
2	B	286	158	158	52±5
3	P	570	554	554	68±8
All	All	22760	17420	17420	2972

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 74.

5 of 606 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:7:DG:H2''	1:A:8:DA:O5'	1.01	1.54	9	13
1:A:9:DT:C6	1:A:10:DT:H72	1.00	1.92	9	20
1:A:7:DG:H2''	1:A:8:DA:C8	0.96	1.94	19	19
2:B:18:DC:H1'	2:B:19:DA:O4'	0.92	1.62	4	1
3:P:56:TYR:O	3:P:60:ILE:HD13	0.92	1.65	11	2

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	P	68/82 (83%)	54±2 (79±2%)	8±2 (12±2%)	6±1 (9±1%)	1	12
All	All	1360/1640 (83%)	1070 (79%)	170 (12%)	120 (9%)	1	12

5 of 10 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
3	P	42	VAL	20
3	P	60	ILE	20
3	P	7	ASN	19
3	P	6	ARG	18
3	P	21	TYR	16

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	P	61/70 (87%)	55±1 (91±2%)	6±1 (9±2%)	13	59
All	All	1220/1400 (87%)	1109 (91%)	111 (9%)	13	59

5 of 19 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	P	36	LYS	20
3	P	47	ASN	18
3	P	31	LYS	17
3	P	58	LYS	13
3	P	10	LYS	7

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

No chemical shift data were provided