wwPDB NMR Structure Validation Summary Report

Oct 17, 2021 – 09:24 AM EDT

PDB ID	:	1I4V
Title	:	SOLUTION STRUCTURE OF THE UMUD' HOMODIMER
Authors	:	Ferentz, A.E.; Walker, G.C.; Wagner, G.
Deposited on	:	2001-02-23

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:

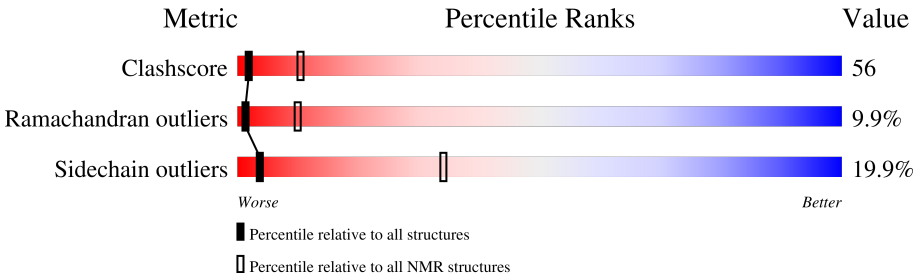
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.23.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

1 Overall quality at a glance (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	115	28% 48% 12% • 11%
1	B	115	26% 50% 12% • 11%

2 Ensemble composition and analysis

This entry contains 20 models. Model 6 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average*.

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:38-A:139, B:38-B:139 (204)	0.79	6

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 3 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 3474 atoms, of which 1744 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called UMUD' PROTEIN.

Mol	Chain	Residues	Atoms						Trace
1	A	115	Total	C	H	N	O	S	0
			1737	550	872	141	171	3	
1	B	115	Total	C	H	N	O	S	0
			1737	550	872	141	171	3	

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	25	ALA	GLY	engineered mutation	UNP P04153
B	25	ALA	GLY	engineered mutation	UNP P04153

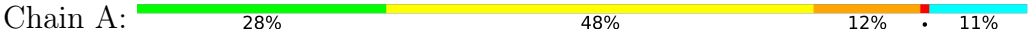
4 Residue-property plots

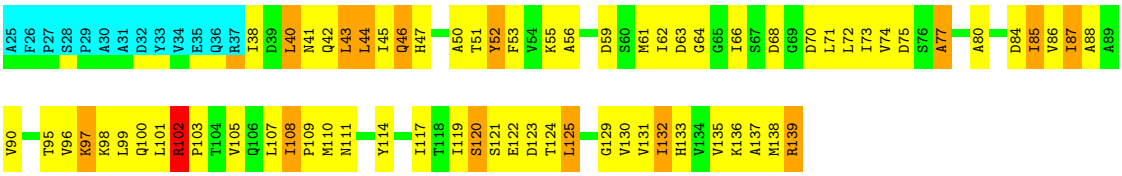
4.1 Average score per residue in the NMR ensemble

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: UMUD' PROTEIN

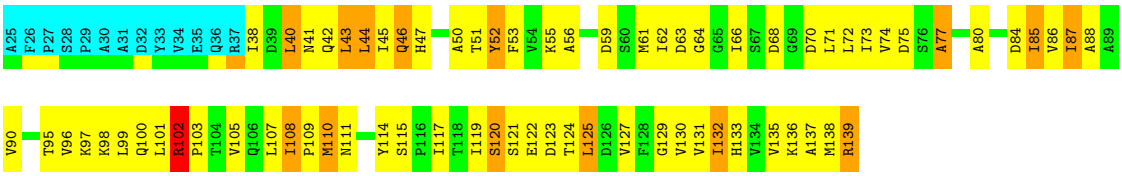
Chain A:





• Molecule 1: UMUD' PROTEIN

Chain B:

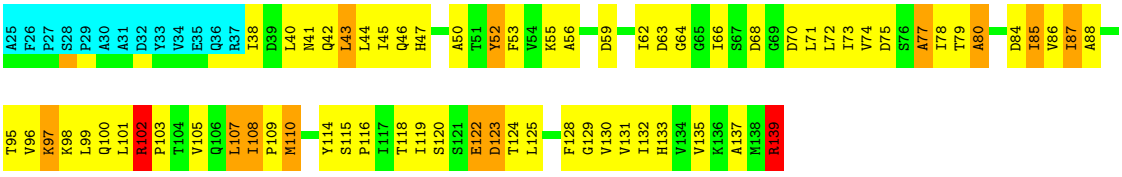


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 6. Colouring as in section 4.1 above.

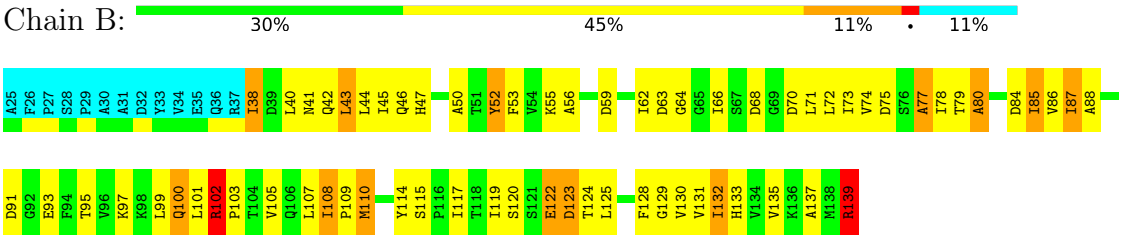
• Molecule 1: UMUD' PROTEIN

Chain A:



• Molecule 1: UMUD' PROTEIN

Chain B:


30%
45%
11%
11%
A25 F26 P27 S28 P29 A30 A31 D32 Y33 V34 E35 Q36 R37 T38 D39 L40 N41 Q42 L43 L44 I45 Q46 H47 A50 T51 Y52 F53 V54 K55 A56 D59 I62 D63 G64 G65 I66 S67 D68 G69 D70 L71 L72 I73 V74 D75 S76 A77 I78 T79 A80 D84 I85 V86 I87 A88
D91 G92 E93 F94 T95 V96 K97 K98 L99 Q100 L101 R102 P103 T104 V105 Q106 L107 I108 P109 M110 Y114 S115 P116 I117 T118 I119 S120 S121 E122 D123 T124 L125 F128 G129 V130 V131 I132 H133 V134 V135 K136 A137 M138 R139

5 Refinement protocol and experimental data overview

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

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

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

Software name	Classification	Version
X-PLOR	refinement	3.851

No chemical shift data was provided.

6 Model quality

6.1 Standard geometry

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
1	A	0.0±0.0	2.0±0.0
1	B	0.0±0.0	2.0±0.0
All	All	0	80

There are no bond-length outliers.

There are no bond-angle outliers.

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	A	102	ARG	Sidechain	20
1	A	139	ARG	Sidechain	20
1	B	102	ARG	Sidechain	20
1	B	139	ARG	Sidechain	20

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	763	778	778	100±9
1	B	763	778	778	101±11
All	All	30520	31120	31120	3481

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 56.

5 of 1113 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:45:ILE:HD11	1:B:40:LEU:HD13	1.05	1.27	11	3
1:A:44:LEU:HD23	1:B:44:LEU:HD23	1.04	1.20	8	12
1:B:107:LEU:HD13	1:B:119:ILE:HD11	1.01	1.26	11	2
1:A:40:LEU:HD13	1:B:45:ILE:HD11	1.00	1.29	11	3
1:B:95:THR:HG21	1:B:107:LEU:HD23	1.00	1.29	9	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
1	A	101/115 (88%)	70±3 (69±3%)	21±3 (21±3%)	10±2 (10±2%)	1 11
1	B	101/115 (88%)	70±3 (69±3%)	21±3 (21±3%)	10±3 (10±3%)	1 9
All	All	4040/4600 (88%)	2792 (69%)	849 (21%)	399 (10%)	1 10

5 of 61 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	77	ALA	20
1	B	77	ALA	20
1	A	80	ALA	13
1	B	80	ALA	13
1	A	68	ASP	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	
1	A	87/97 (90%)	69±3 (80±3%)	18±3 (20±3%)	3	33
1	B	87/97 (90%)	70±4 (80±4%)	17±4 (20±4%)	4	34
All	All	3480/3880 (90%)	2786 (80%)	694 (20%)	4	34

5 of 101 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	43	LEU	20
1	A	55	LYS	20
1	B	43	LEU	20
1	B	55	LYS	20
1	A	52	TYR	17

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