wwPDB NMR Structure Validation Summary Report

Jun 12, 2024 – 11:11 AM EDT

PDB ID	:	2KOX
Title	:	NMR residual dipolar couplings identify long range correlated motions in the backbone of the protein ubiquitin
Authors	:	Fenwick, R.B.; Richter, B.; Lee, D.; Walter, K.F.A.; Milovanovic, D.; Becker, S.; Lakomek, N.A.; Griesinger, C.; Salvatella, X.
Deposited on	:	2009-10-02

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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

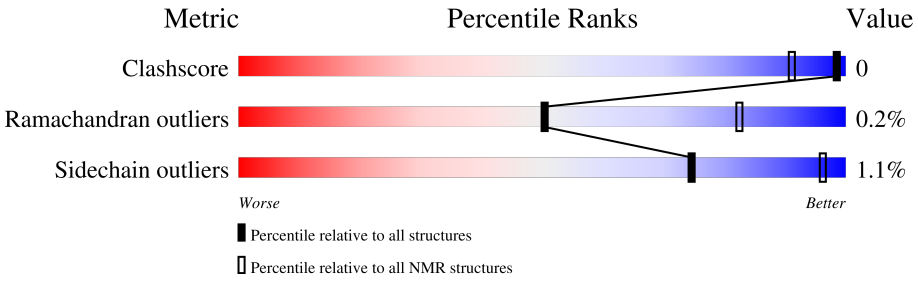
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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	76	89% · 7%

2 Ensemble composition and analysis (i)

This entry contains 640 models. Model 76 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:1-A:71 (71)	0.60	76

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 79 clusters and 27 single-model clusters were found.

Cluster number	Models
1	1, 39, 57, 61, 101, 103, 121, 144, 150, 151, 158, 163, 165, 167, 180, 184, 196, 213, 231, 233, 235, 238, 247, 251, 252, 253, 255, 262, 264, 268, 288, 293, 297, 315, 319, 330, 332, 333, 341, 366, 377, 379, 380, 405, 415, 423, 460, 469, 483, 507, 512, 547, 559, 564, 571, 572, 592, 603, 610
2	2, 24, 26, 33, 37, 47, 49, 62, 74, 81, 88, 111, 129, 154, 169, 175, 179, 204, 218, 256, 258, 276, 280, 282, 286, 303, 311, 320, 322, 335, 344, 350, 361, 363, 365, 367, 371, 384, 394, 404, 408, 427, 429, 435, 448, 452, 468, 470, 474, 529, 532, 534, 555, 567, 578, 596, 600, 602
3	8, 11, 21, 22, 31, 52, 54, 73, 106, 108, 116, 138, 149, 182, 189, 211, 214, 216, 246, 275, 342, 364, 417, 428, 440, 442, 445, 447, 472, 492, 504, 506, 509, 511, 519, 533, 566, 573, 583, 625, 631
4	12, 38, 40, 66, 69, 76, 78, 79, 86, 97, 102, 104, 142, 148, 161, 166, 200, 212, 230, 236, 245, 277, 300, 309, 339, 347, 358, 373, 374, 396, 426, 437, 449, 462, 481, 501, 524, 526, 554, 590, 626
5	27, 65, 159, 195, 234, 257, 287, 298, 325, 362, 383, 387, 389, 399, 424, 453, 461, 486, 488, 515, 517, 525, 550, 570, 579, 633
6	29, 48, 50, 60, 89, 91, 210, 217, 225, 227, 242, 281, 289, 321, 345, 355, 370, 432, 434, 496, 601
7	28, 117, 181, 220, 222, 278, 284, 292, 294, 348, 369, 410, 412, 458, 478, 538, 540, 588, 604, 613
8	16, 80, 92, 208, 274, 334, 398, 400, 457, 464, 466, 499, 530, 539, 560, 594, 623, 628
9	19, 23, 174, 190, 299, 316, 326, 329, 390, 403, 436, 454, 471, 518, 535, 561, 582

10	15, 99, 171, 215, 224, 237, 243, 279, 343, 407, 446, 476, 542, 574
11	30, 32, 94, 202, 266, 290, 418, 494, 522, 536, 544, 586, 609
12	93, 157, 176, 240, 285, 381, 402, 443, 562, 589, 605
13	9, 55, 59, 123, 125, 187, 191, 198, 441, 479, 599
14	14, 46, 115, 119, 183, 324, 414, 516, 597, 607, 632
15	13, 77, 114, 127, 141, 155, 178, 205, 306, 615, 634
16	44, 120, 130, 226, 267, 395, 521, 568, 581, 587
17	41, 64, 105, 192, 241, 254, 318, 508, 563, 618
18	25, 139, 239, 397, 455, 463, 503, 528, 617, 622
19	34, 98, 223, 351, 413, 477, 541, 619
20	5, 20, 43, 135, 168, 201, 259, 263
21	17, 96, 145, 209, 354, 386, 433, 465
22	56, 72, 110, 118, 248, 250, 378, 430
23	229, 357, 421, 489, 502, 553, 616
24	100, 134, 136, 170, 244, 495, 575
25	10, 513, 577, 595, 612, 640
26	90, 109, 301, 425, 491, 510
27	18, 71, 112, 199, 323, 451
28	85, 438, 556, 621, 624, 629
29	147, 328, 338, 392, 498, 580
30	295, 487, 551, 614, 638
31	68, 188, 375, 485, 549
32	107, 382, 450, 514, 639
33	143, 207, 271, 591
34	261, 393, 411, 475
35	3, 193, 490, 537
36	51, 132, 260, 576
37	133, 523, 565, 585
38	75, 439, 505, 636
39	356, 420, 484, 611
40	137, 269, 331, 598
41	83, 467, 531
42	6, 337, 401
43	352, 520, 584
44	177, 314, 497
45	58, 327, 391
46	221, 368, 543
47	340, 422, 569
48	95, 146, 630
49	265, 310, 593
50	87, 353, 409
51	70, 273
52	232, 552

53	482, 546
54	500, 627
55	128, 194
56	173, 493
57	186, 444
58	307, 637
59	473, 635
60	346, 388
61	122, 527
62	160, 406
63	164, 548
64	419, 606
65	67, 296
66	308, 372
67	203, 459
68	152, 317
69	249, 313
70	63, 219
71	124, 385
72	45, 126
73	270, 557
74	42, 185
75	172, 558
76	156, 431
77	140, 456
78	416, 480
79	283, 336
Single-model clusters	4; 7; 35; 36; 53; 82; 84; 113; 131; 153; 162; 197; 206; 228; 272; 291; 302; 304; 305; 31

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1231 atoms, of which 629 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Ubiquitin.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	76	1231	378	629	105	118	1	0

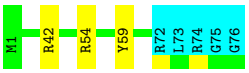
4 Residue-property plots

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: Ubiquitin

Chain A:



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

The representative model is number 76. Colouring as in section 4.1 above.

- Molecule 1: Ubiquitin

Chain A:

5 Refinement protocol and experimental data overview

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

Of the 640 calculated structures, 640 were deposited, based on the following criterion: *all calculated structures submitted.*

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

Software name	Classification	Version
CHARMM	structure solution	c34b1
CHARMM	refinement	c34b1

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	A	1.37±0.05	1±1/570 (0.2± 0.2%)	1.70±0.07	7±3/770 (0.9± 0.3%)
All	All	1.37	655/364800 (0.2%)	1.70	4559/492800 (0.9%)

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	0.8±0.8
All	All	0	538

5 of 116 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
1	A	16	GLU	CG-CD	10.83	1.68	1.51	414	10
1	A	57	SER	CA-CB	8.60	1.65	1.52	590	32
1	A	65	SER	CB-OG	8.10	1.52	1.42	439	4
1	A	59	TYR	CG-CD1	8.06	1.49	1.39	151	21
1	A	42	ARG	CD-NE	8.01	1.60	1.46	551	27

5 of 311 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(°)	Ideal(°)	Models	
								Worst	Total
1	A	42	ARG	NE-CZ-NH1	20.09	130.34	120.30	239	326
1	A	54	ARG	NE-CZ-NH2	-19.78	110.41	120.30	32	305
1	A	54	ARG	NE-CZ-NH1	18.53	129.57	120.30	428	336
1	A	42	ARG	NE-CZ-NH2	-18.11	111.24	120.30	544	235
1	A	59	TYR	CB-CG-CD1	-16.69	110.98	121.00	595	178

There are no chirality outliers.

5 of 20 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	42	ARG	Sidechain	171
1	A	54	ARG	Sidechain	157
1	A	59	TYR	Sidechain	121
1	A	45	PHE	Sidechain	25
1	A	4	PHE	Sidechain	18

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	563	586	586	0±0
All	All	360320	375040	375040	61

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 0.

5 of 41 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:3:ILE:HD12	1:A:67:LEU:HD22	0.59	1.74	483	1
1:A:23:ILE:HG23	1:A:43:LEU:HD12	0.54	1.77	267	2
1:A:22:THR:O	1:A:26:VAL:HG23	0.53	2.04	359	3
1:A:30:ILE:HG21	1:A:69:LEU:HD22	0.51	1.83	30	1
1:A:23:ILE:HD13	1:A:56:LEU:HD12	0.50	1.82	533	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	70/76 (92%)	45±32 (65±46%)	1±1 (1±2%)	0±0 (0±0%)	50 82
All	All	29610/48640 (61%)	28912 (98%)	627 (2%)	71 (0%)	50 82

5 of 15 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	63	LYS	17
1	A	11	LYS	9
1	A	37	PRO	9
1	A	52	ASP	9
1	A	50	LEU	8

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	65/68 (96%)	64±1 (99±1%)	1±1 (1±1%)	74 96
All	All	41600/43520 (96%)	41130 (99%)	470 (1%)	74 96

5 of 50 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	52	ASP	57
1	A	38	PRO	49
1	A	1	MET	39
1	A	19	PRO	21
1	A	17	VAL	19

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