wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 04:04 PM BST

PDB ID	:	1JOQ
Title	:	Ensemble structures for Staphylococcal nuclease-H124L in ternary complex with Ca2+ and thymidine-3',5'-bisphosphate
Authors	:	Wang, J.; Truckses, D.M.; Abildgaard, F.; Dzakula, Z.; Zolnai, Z.; Markley, J.L.
Deposited on	:	2001-07-30

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
http://wwpdb.org/validation/2016/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:

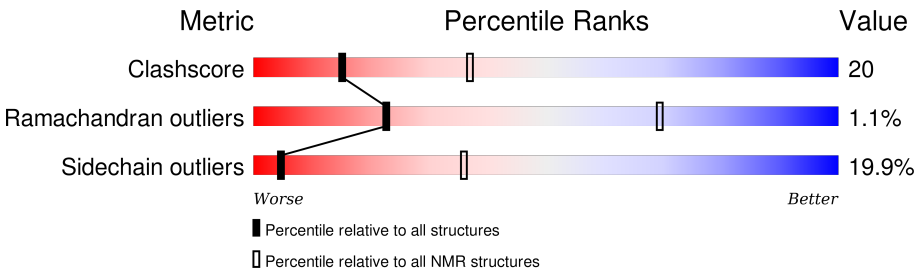
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027457

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	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

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	149	42% 34% • 21%

2 Ensemble composition and analysis (i)

This entry contains 30 models. Model 14 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:8-A:41, A:57-A:140 (118)	0.31	14

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 6 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2432 atoms, of which 1226 are hydrogens and 0 are deuteriums.

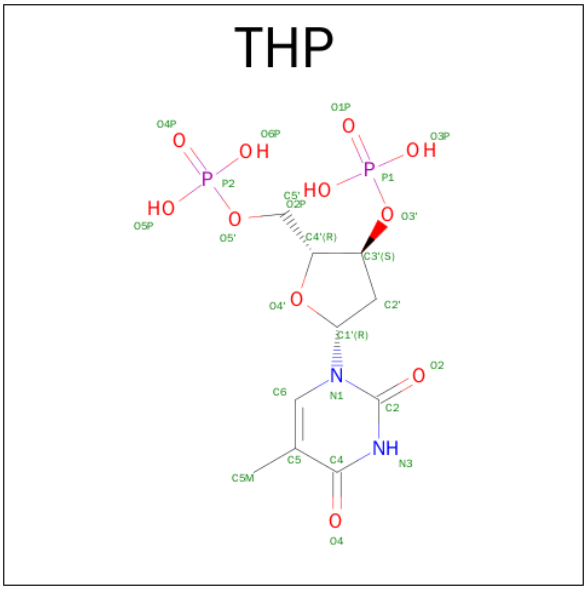
- Molecule 1 is a protein called staphylococcal nuclease.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	149	2397	747	1216	206	224	4	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	124	LEU	HIS	ENGINEERED	UNP P00644

- Molecule 2 is THYMIDINE-3',5'-DIPHOSPHATE (three-letter code: THP) (formula: $C_{10}H_{16}N_2O_{11}P_2$).



Mol	Chain	Residues	Atoms					
			Total	C	H	N	O	P
2	A	1	35	10	10	2	11	2

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: staphylococcal nuclease

Chain A: 42% 34% • 21%

A1 T2 S3 T4 K5 K6 L7 H8 K9 E10 T13 L14 I18 D21 T22 V23 K24 L25 M26 Y27 K28 T33 F34 R35 L36 L37 L38 V39 D40 T41 P42 E43 T44 K45 H46 P47 K48 K49 G50 V51 E52 K53 Y54 G55 P56 E57 A58 S59 T62 K63 K64 M65 V66 E67

K78 R81 T82 D83 K84 Y85 G86 R87 G88 L89 A94 D95 G96 K97 N100 L103 V104 R105 Q106 G107 E122 R126 K127 S128 E129 K133 K134 E135 K136 L137 N138 I139 W140 S141 E142 D143 N144 A145 D146 S147 G148 Q149

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

The representative model is number 14. Colouring as in section 4.1 above.

- Molecule 1: staphylococcal nuclease

Chain A: 44% 32% • 21%

A1 T2 S3 T4 K5 K6 L7 H8 K9 L14 I18 D19 G20 D21 T22 V23 K24 L25 T33 F34 R35 L36 L37 L38 V39 D40 T41 P42 E43 T44 K45 H46 P47 K48 K49 G50 V51 E52 K53 Y54 G55 P56 E57 A58 S59 A60 F61 T62 K63 K64 M65 V66 K70 E75

K78 G79 Q80 R81 K84 Y85 G86 R87 G88 L89 V99 N100 L103 V104 R105 Q106 G107 L108 E122 Q123 L124 L125 R126 K127 S128 E129 K133 K134 E135 K136 L137 N138 I139 W140 S141 E142 D143 N144 A145 D146 S147 G148 Q149

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *Distance, hydrogen bond and torsion angle constraints.*

Of the 90 calculated structures, 30 were deposited, based on the following criterion: *structures with acceptable covalent geometry,structures with the least restraint violations.*

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

Software name	Classification	Version
X-PLOR	structure solution	3.1
INSIGHTII	structure solution	2.3
HBPLUS	structure solution	
X-PLOR	refinement	3.1

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

6 Model quality (i)

6.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: THP

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	4.6±0.5
All	All	0	139

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	81	ARG	Sidechain	29
1	A	126	ARG	Sidechain	29
1	A	87	ARG	Sidechain	29
1	A	105	ARG	Sidechain	28
1	A	35	ARG	Sidechain	24

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	947	984	984	38±6
2	A	25	10	12	4±3
All	All	29160	29820	29880	1152

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 20.

5 of 496 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:62:THR:O	1:A:66:VAL:HG22	0.93	1.64	25	27
1:A:21:ASP:OD1	1:A:36:LEU:HD12	0.89	1.67	17	5
1:A:107:GLY:O	1:A:139:ILE:HD12	0.89	1.66	7	22
1:A:14:LEU:HD13	1:A:15:ILE:N	0.88	1.83	7	2
1:A:36:LEU:HD22	1:A:100:ASN:OD1	0.87	1.69	17	17

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	118/149 (79%)	102±3 (86±2%)	15±3 (13±2%)	1±1 (1±1%)	23	69
All	All	3540/4470 (79%)	3050 (86%)	451 (13%)	39 (1%)	23	69

5 of 12 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	138	ASN	13
1	A	57	GLU	6
1	A	65	MET	4
1	A	139	ILE	3
1	A	113	TYR	3

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	99/125 (79%)	79±4 (80±4%)	20±4 (20±4%)	5 36
All	All	2970/3750 (79%)	2379 (80%)	591 (20%)	5 36

5 of 64 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	78	LYS	30
1	A	137	LEU	23
1	A	97	LYS	21
1	A	64	LYS	20
1	A	65	MET	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 carbohydrates in this entry.

6.6 Ligand geometry (i)

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
2	THP	A	150	-	19,26,26	1.14±0.01	0±0 (0±0%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of angles that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
2	THP	A	150	-	28,40,40	1.06±0.01	0±0 (0±0%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	THP	A	150	-	-	0±0,11,27,27	0±0,2,2,2

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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