wwPDB NMR Structure Validation Summary Report (i)

Oct 23, 2021 – 05:56 PM EDT

PDB ID	:	1AOU
Title	:	NMR STRUCTURE OF THE FYN SH2 DOMAIN COMPLEXED WITH A PHOSPHOTYROSYL PEPTIDE, 22 STRUCTURES
Authors	:	Mulhern, T.D.; Shaw, G.L.; Morton, C.J.; Day, A.J.; Campbell, I.D.
Deposited on	:	1997-07-10

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 (1)) were used in the production of this report:

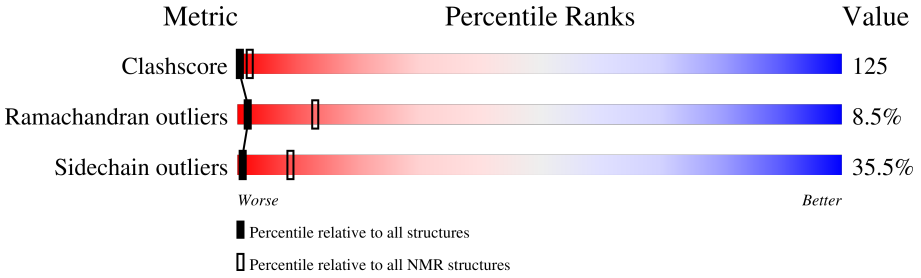
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.23.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal 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	F	106	8% 58% 24% 6% 5%
2	P	11	9% 18% 73%

2 Ensemble composition and analysis (i)

This entry contains 22 models. Model 18 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	F:2-F:102, P:205-P:207 (104)	0.51	18

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, 3, 4, 5, 6, 8, 9, 10, 11, 12, 14, 15, 17, 18, 20, 22
2	7, 16, 19
3	1, 13, 21

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1918 atoms, of which 946 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called FYN PROTEIN-TYROSINE KINASE.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	F	106	1724	547	855	159	162	1	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	97	SER	CYS	engineered mutation	UNP P06241
F	98	SER	CYS	engineered mutation	UNP P06241
F	104	SER	CYS	engineered mutation	UNP P06241

- Molecule 2 is a protein called PHOSPHOTYROSYL PEPTIDE.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	P	11	194	66	91	12	24	1	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: FYN PROTEIN-TYROSINE KINASE

Chain F: 8% 58% 24% 6% 5%

S1 I2 Q3 A4 E5 E6 W7 Y8 F9 G10 K11 L12 G13 R14 K15 D16 A17 E18 R19 Q20 L21 L22 S23 F24 G25 N26 P27 R28 G29 T30 F31 L32 I33 R34 E35 S36 E37 T38 T39 K40 G41 A42 Y43 S44 L45 S46 I47 R48 D49 W50 D51 D52 M53 K54 G55 D56 H57 V58 K59 H60

Y61 K62 I63 R64 K65 L66 D67 G70 Y71 Y72 I73 T74 T75 R76 A77 Q78 F79 E80 T81 L82 Q83 Q84 L85 V86 Q87 H88 Y89 S90 E91 R92 A93 A94 G95 L96 S97 S98 R99 L100 V101 V102 P103 S104 H105 K106

• Molecule 2: PHOSPHOTYROSYL PEPTIDE

Chain P: 9% 18% 73%

E201 P202 Q203 Y204 E205 E206 I207 P208 I209 Y210 L211

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

The representative model is number 18. Colouring as in section 4.1 above.

• Molecule 1: FYN PROTEIN-TYROSINE KINASE

Chain F: 14% 53% 24% 5% 5%

S1 I2 Q3 A4 E5 E6 W7 Y8 F9 G10 K11 L12 G13 R14 K15 D16 A17 E18 R19 Q20 L21 L22 S23 F24 G25 N26 P27 R28 G29 T30 F31 L32 I33 R34 E35 S36 E37 T38 T39 K40 G41 A42 Y43 S44 L45 S46 I47 R48 D49 W50 D51 D52 M53 K54 G55 D56 H57 V58 K59 H60

Y61 K62 I63 R64 K65 L66 D67 G70 Y71 Y72 I73 T74 T75 R76 A77 Q78 F79 E80 T81 L82 Q83 Q84 L85 V86 Y89 S90 E91 R92 A93 A94 G95 L96 S97 S98 R99 L100 V101 V102 P103 S104 H105 K106

• Molecule 2: PHOSPHOTYROSYL PEPTIDE

Chain P: 9% 18% 73%

E201
P202
Q203
Y204
E205
E206
I207
P208
I209
Y210
L211

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *DYNAMICAL SIMULATED ANNEALING*.

Of the 200 calculated structures, 22 were deposited, based on the following criterion: *LEAST VIOLATIONS*.

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

Software name	Classification	Version
X-PLOR	refinement	3.1
X-PLOR	structure solution	

No chemical shift data was provided.

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: PTR

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	F	0.0±0.0	8.5±0.6
All	All	0	188

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

5 of 9 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	F	28	ARG	Sidechain	22
1	F	92	ARG	Sidechain	22
1	F	99	ARG	Sidechain	22
1	F	14	ARG	Sidechain	21
1	F	19	ARG	Sidechain	21

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	F	830	816	816	212±14
2	P	26	23	23	5±3
All	All	18832	18458	18458	4675

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 125.

5 of 1317 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:F:7:TRP:CZ2	1:F:86:VAL:HG11	1.07	1.84	9	8
1:F:2:ILE:HD11	1:F:8:TYR:CG	1.05	1.87	6	7
1:F:7:TRP:CD2	1:F:86:VAL:HG11	1.03	1.87	21	16
1:F:21:LEU:HD12	1:F:58:VAL:HG22	1.02	1.31	12	1
1:F:21:LEU:HD21	1:F:32:LEU:HD23	1.01	1.30	17	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
1	F	101/106 (95%)	71±3 (70±3%)	21±3 (21±3%)	9±2 (9±2%)	1 12
2	P	3/11 (27%)	3±0 (88±16%)	0±0 (12±16%)	0±0 (0±0%)	100 100
All	All	2288/2574 (89%)	1622 (71%)	471 (21%)	195 (9%)	2 13

5 of 29 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	F	41	GLY	22
1	F	54	LYS	18
1	F	5	GLU	16
1	F	99	ARG	16
1	F	100	LEU	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	F	86/91 (95%)	55±3 (64±4%)	31±3 (36±4%)	1 8
2	P	3/10 (30%)	3±0 (91±15%)	0±0 (9±15%)	13 59
All	All	1958/2222 (88%)	1263 (65%)	695 (35%)	1 9

5 of 71 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	F	14	ARG	22
1	F	43	TYR	22
1	F	74	THR	22
1	F	81	THR	22
1	F	85	LEU	22

6.3.3 RNA (i)

There are no RNA molecules in this entry.

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

1 non-standard protein/DNA/RNA residue 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	PTR	P	204	2	15,16,17	0.94±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	PTR	P	204	2	19,22,24	0.89±0.01	1±0 (5±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	PTR	P	204	2	-	0±0,10,11,13	0±0,1,1,1

There are no bond-length outliers.

All unique angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
2	P	204	PTR	CG-CB-CA	2.35	109.35	114.10	11	22

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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