wwPDB NMR Structure Validation Summary Report

Apr 20, 2024 – 03:56 PM EDT

PDB ID	:	2JXC
Title	:	Structure of the EPS15-EH2 Stonin2 Complex
Authors	:	Rumpf, J.; Simon, B.; Jung, N.; Maritzen, T.; Haucke, V.; Sattler, M.; Groemping, Y.
Deposited on	:	2007-11-12

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 (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:

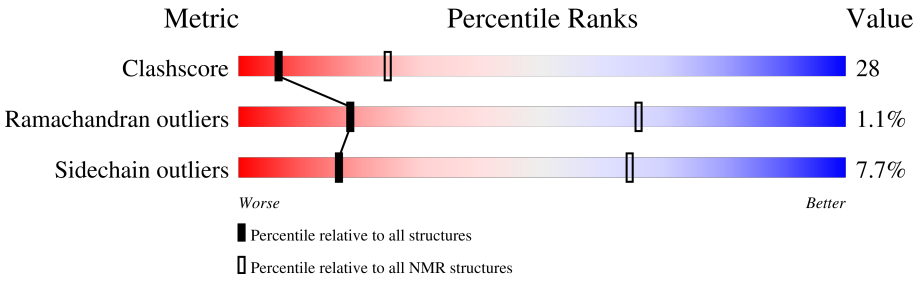
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson 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	100	44% 46% • • 5%
2	B	45	11% 27% 51% 11%

2 Ensemble composition and analysis

This entry contains 10 models. Model 7 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	A:121-A:212, B:309-B:317, B:325-B:332 (109)	0.47	7

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 and 3 single-model clusters were found.

Cluster number	Models
1	3, 5, 6
2	7, 9
3	1, 2
Single-model clusters	4; 8; 10

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 2119 atoms, of which 1050 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Epidermal growth factor receptor substrate 15.

Mol	Chain	Residues	Atoms						Trace
1	A	95	Total	C	H	N	O	S	0
			1496	484	753	118	137	4	

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	116	GLY	-	expression tag	UNP P42566
A	117	PRO	-	expression tag	UNP P42566
A	118	LEU	-	expression tag	UNP P42566
A	119	GLY	-	expression tag	UNP P42566
A	120	SER	-	expression tag	UNP P42566

- Molecule 2 is a protein called Stonin-2.

Mol	Chain	Residues	Atoms					Trace
2	B	40	Total	C	H	N	O	0
			622	206	297	53	66	

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	296	GLY	-	expression tag	UNP Q8WXE9
B	297	PRO	-	expression tag	UNP Q8WXE9
B	298	LEU	-	expression tag	UNP Q8WXE9
B	299	GLY	-	expression tag	UNP Q8WXE9
B	300	SER	-	expression tag	UNP Q8WXE9

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	
3	A	1	Total	Ca
			1	1

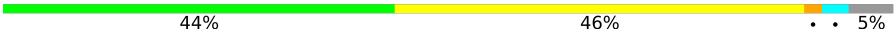
4 Residue-property plots

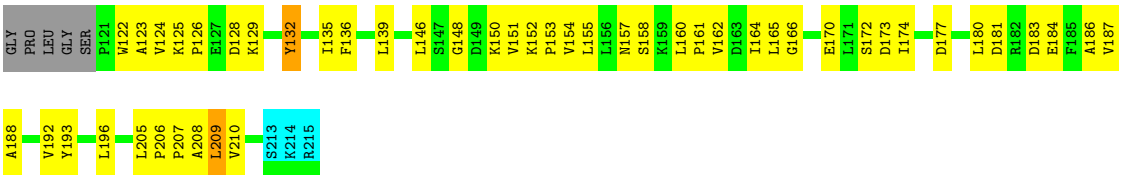
4.1 Average score per residue in the NMR ensemble

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: Epidermal growth factor receptor substrate 15

Chain A:





• Molecule 2: Stonin-2

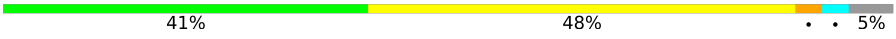
Chain B:

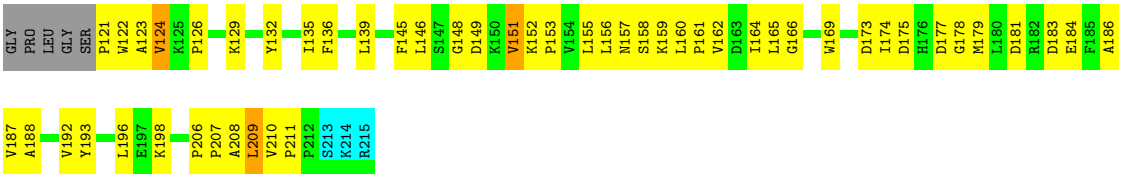
4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 7. Colouring as in section 4.1 above.

• Molecule 1: Epidermal growth factor receptor substrate 15

Chain A:





• Molecule 2: Stonin-2

Chain B:


GLY
PRO
LEU
GLY
SER
P301
S302
V303
T304
E305
A306
S307
P308
W309
R310
A311
T312
N313
P314
N317
E318
T319
L320
Q321
D322
V323
Q324
P325
S326
P327
I328
N329
P330
F331
S332
A333
F334
F335
E336
E337
Q338
E339
R340

5 Refinement protocol and experimental data overview

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

Of the 100 calculated structures, 10 were deposited, based on the following criterion: *structures with the lowest energy.*

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

Software name	Classification	Version
CNS	refinement	1.1
ARIA	structure solution	1.2

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: CA

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	0.55±0.13	2±1/738 (0.2± 0.2%)	0.57±0.03	0±0/1009 (0.0± 0.0%)
2	B	0.47±0.08	0±0/147 (0.0± 0.0%)	0.56±0.03	0±0/204 (0.0± 0.0%)
All	All	0.55	16/8850 (0.2%)	0.57	0/12130 (0.0%)

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	0.2±0.4
All	All	0	2

All 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	132	TYR	CE2-CZ	10.36	1.52	1.38	8	4
1	A	132	TYR	CE1-CZ	-9.98	1.25	1.38	8	4
1	A	193	TYR	CE1-CZ	-6.63	1.29	1.38	9	4
1	A	193	TYR	CE2-CZ	6.59	1.47	1.38	9	4

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	193	TYR	Sidechain	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group	Models (Total)
1	A	132	TYR	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	716	722	721	44±4
2	B	139	131	131	14±5
3	A	1	0	0	2±0
All	All	8560	8530	8520	478

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 28.

5 of 187 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:160:LEU:HB3	1:A:164:ILE:HD11	1.04	1.28	2	9
1:A:155:LEU:HB3	1:A:165:LEU:HD22	0.92	1.39	5	10
1:A:124:VAL:HG11	1:A:186:ALA:HB1	0.91	1.41	1	10
1:A:206:PRO:HD2	1:A:209:LEU:HD11	0.88	1.43	8	1
1:A:157:ASN:HB2	2:B:330:PRO:HD2	0.79	1.53	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	91/100 (91%)	86±1 (94±1%)	4±1 (5±2%)	1±1 (1±1%)	20	68
2	B	17/45 (38%)	11±1 (65±8%)	6±1 (33±8%)	0±1 (2±4%)	12	54

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	1080/1450 (74%)	969 (90%)	99 (9%)	12 (1%)	18 66

All 4 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	126	PRO	8
2	B	328	ILE	2
2	B	325	PRO	1
1	A	123	ALA	1

6.3.2 Protein sidechains

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	81/87 (93%)	75±2 (93±2%)	6±2 (7±2%)	19 67
2	B	16/40 (40%)	14±1 (89±9%)	2±1 (11±9%)	9 53
All	All	970/1270 (76%)	895 (92%)	75 (8%)	16 64

5 of 29 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	209	LEU	10
1	A	125	LYS	7
2	B	310	ARG	6
1	A	152	LYS	5
1	A	149	ASP	5

6.3.3 RNA

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)

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

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