wwPDB NMR Structure Validation Summary Report (i)

Feb 23, 2022 – 07:56 AM EST

PDB ID	:	1X5A
Title	:	The solution structure of the second fibronectin type III domain of mouse Ephrin type-A receptor 1
Authors	:	Tochio, N.; Sasagawa, A.; Koshiba, S.; Inoue, M.; Kigawa, T.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2005-05-15

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:

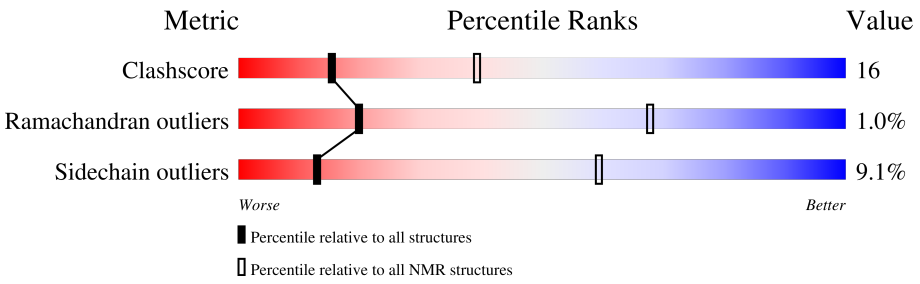
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.26
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.26

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	107	53% 17% • 28%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 1 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:15-A:31, A:42-A:101 (77)	0.26	1

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 4 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1640 atoms, of which 818 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Ephrin type-A receptor 1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	107	1640	515	818	147	159	1	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	cloning artifact	UNP Q60750
A	2	SER	-	cloning artifact	UNP Q60750
A	3	SER	-	cloning artifact	UNP Q60750
A	4	GLY	-	cloning artifact	UNP Q60750
A	5	SER	-	cloning artifact	UNP Q60750
A	6	SER	-	cloning artifact	UNP Q60750
A	7	GLY	-	cloning artifact	UNP Q60750
A	102	SER	-	cloning artifact	UNP Q60750
A	103	GLY	-	cloning artifact	UNP Q60750
A	104	PRO	-	cloning artifact	UNP Q60750
A	105	SER	-	cloning artifact	UNP Q60750
A	106	SER	-	cloning artifact	UNP Q60750
A	107	GLY	-	cloning artifact	UNP Q60750

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: Ephrin type-A receptor 1

Chain A: 53% 17% • 28%

G1 S2 S3 G4 S5 S6 G7 A8 E9 S10 L11 S12 G13 L14 S15 L16 V19 K20 L26 E27 L28 T29 W30 A31 G32 S33 R34 P35 R36 N37 P38 G39 G40 N41 L46 H47 V48 L49 N50 Q51 H56 L65 V77 R78 V79 R80 L85 H94 E95 F96 S102

G103 P104 S105 S106 G107

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

The representative model is number 1. Colouring as in section 4.1 above.

• Molecule 1: Ephrin type-A receptor 1

Chain A: 48% 22% • 28%

G1 S2 S3 G4 S5 S6 G7 A8 E9 S10 L11 S12 G13 L14 S15 L16 K17 L26 E27 L28 T29 W30 A31 G32 S33 R34 P35 R36 N37 P38 G39 G40 N41 L42 E45 L46 H47 V48 D52 E53 E54 W55 H56 L65 D72 V77 R78 V79 R80 T81 L82 F90

H94 E95 F96 R97 T98 S102 G103 P104 S105 S106 G107

5 Refinement protocol and experimental data overview (i)

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

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

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

Software name	Classification	Version
CYANA	structure solution	2.0.17
CYANA	refinement	2.0.17

No chemical shift data was provided.

6 Model quality (i)

6.1 Standard geometry (i)

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	637	649	646	21±5
All	All	12740	12980	12920	423

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 16.

5 of 135 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:30:TRP:CZ2	1:A:79:VAL:HG21	1.09	1.83	17	20
1:A:16:LEU:HD22	1:A:28:LEU:HD13	0.94	1.39	17	3
1:A:16:LEU:HD12	1:A:94:HIS:CD2	0.92	1.98	18	1
1:A:26:LEU:HD11	1:A:98:THR:HG22	0.81	1.52	13	7
1:A:30:TRP:CE2	1:A:79:VAL:HG21	0.74	2.16	19	18

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	77/107 (72%)	72±2 (93±2%)	5±1 (6±2%)	1±1 (1±1%)	20	68
All	All	1540/2140 (72%)	1434 (93%)	91 (6%)	15 (1%)	20	68

5 of 6 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	93	ASP	5
1	A	86	GLY	4
1	A	51	GLN	2
1	A	16	LEU	2
1	A	42	LEU	1

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	74/94 (79%)	67±2 (91±2%)	7±2 (9±2%)	13	59
All	All	1480/1880 (79%)	1345 (91%)	135 (9%)	13	59

5 of 24 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	46	LEU	20
1	A	56	HIS	20
1	A	65	LEU	15
1	A	80	ARG	11
1	A	50	ASN	9

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