wwPDB NMR Structure Validation Summary Report (i)

Oct 21, 2021 – 08:06 AM EDT

PDB ID	:	1FHS
Title	:	THE THREE-DIMENSIONAL SOLUTION STRUCTURE OF THE SRC HOMOLOGY DOMAIN-2 OF THE GROWTH FACTOR RECEPTOR BOUND PROTEIN-2, NMR, 18 STRUCTURES
Authors	:	Senior, M.M.; Frederick, A.F.; Black, S.; Perkins, L.M.; Wilson, O.; Snow, M.E.; Wang, Y.-S.
Deposited on	:	1997-06-12

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

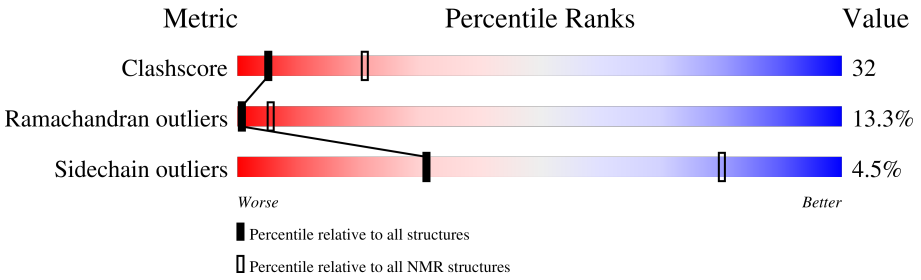
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.23.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (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	A	112	14% 61% • 22%

2 Ensemble composition and analysis (i)

This entry contains 18 models. Model 11 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:61, A:66-A:90, A:94-A:101 (87)	0.59	11

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 2 clusters and 1 single-model cluster was found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1820 atoms, of which 901 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called GROWTH FACTOR RECEPTOR BOUND PROTEIN-2.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	112	1820	589	901	163	165	2	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: GROWTH FACTOR RECEPTOR BOUND PROTEIN-2

Chain A: 14% 61% • 22%

G1 I2 E3 M4 K5 P6 H7 P8 W9 F10 F11 G12 K13 I14 P15 R16 A17 K18 A19 E20 E21 M22 L23 S24 K25 Q26 R27 H28 D29 G30 A31 F32 L33 I34 R35 E36 S37 E38 S39 A40 P41 G42 D43 F44 S45 L46 S47 V48 K49 F50 G51 N52 D53 V54 Q55 H56 F57 K58 V59 L60

R61 D62 G63 A64 G65 K66 Y67 F68 L69 W70 V71 V72 K73 F74 N75 S76 L77 N78 E79 L80 Y83 H84 R85 S86 T87 S88 V89 S90 R91 N92 Q93 Q94 I95 F96 I97 R98 D99 I100 E101 Q102 V103 P104 Q105 Q106 P107 T108 Y109 V110 Q111 A112

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

The representative model is number 11. Colouring as in section 4.1 above.

• Molecule 1: GROWTH FACTOR RECEPTOR BOUND PROTEIN-2

Chain A: 29% 43% 5% 22%

G1 I2 E3 M4 K5 P6 H7 P8 F11 G12 K13 I14 P15 R16 A17 K18 A19 E20 E21 M22 L23 S24 K25 H28 D29 G30 A31 F32 L33 E36 S37 E38 D43 F44 S45 L46 S47 V48 K49 F50 G51 N52 D53 V54 Q55 H56 F57 K58 V59 D62 G63 A64 G65 K66 Y67

F68 L69 W70 V71 V72 K73 F74 N75 S76 L77 N78 E79 L80 T87 S88 V89 S90 R91 N92 Q93 Q94 I95 F96 L97 R98 D99 I100 E101 Q102 V103 P104 Q105 Q106 P107 T108 Y109 V110 Q111 A112

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *DG/SA*.

Of the 100 calculated structures, 18 were deposited, based on the following criterion: *?*.

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 calculation	3.1

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.

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	0.1±0.2
All	All	0	1

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	42	GLY	Peptide	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	726	716	716	47±8
All	All	13068	12888	12888	841

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 32.

5 of 424 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:74:PHE:CZ	1:A:80:LEU:HD12	0.81	2.11	2	2
1:A:40:ALA:HB1	1:A:41:PRO:HA	0.80	1.52	8	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:16:ARG:CG	1:A:35:ARG:HH22	0.69	2.01	9	1
1:A:98:ARG:CD	1:A:99:ASP:H	0.67	2.02	18	1
1:A:61:ARG:CZ	1:A:66:LYS:H	0.65	2.04	5	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	87/112 (78%)	56±5 (64±6%)	20±4 (23±5%)	12±2 (13±2%)	1 5
All	All	1566/2016 (78%)	1004 (64%)	353 (23%)	209 (13%)	1 5

5 of 40 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	21	GLU	15
1	A	8	PRO	15
1	A	100	ILE	14
1	A	94	GLN	13
1	A	71	VAL	11

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	79/99 (80%)	75±1 (95±2%)	4±1 (5±2%)	31 80
All	All	1422/1782 (80%)	1358 (95%)	64 (5%)	31 80

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	28	HIS	6
1	A	56	HIS	5
1	A	13	LYS	5
1	A	61	ARG	4
1	A	66	LYS	3

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