W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

Full wwPDB NMR Structure Validation Report (i)

Aug 17, 2022 – 02:10 am BST

PDB ID	:	7PH8
Title	:	Structure of Insulin-like growth factor 1 receptor's transmembrane domain
Authors	:	Bershatsky, Y.V.; Nadezhdin, K.D.; Bocharova, O.V.; Bocharov, E.V.
Deposited on	:	2021-08-16

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	:	2.29
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.29

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 96%.

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.
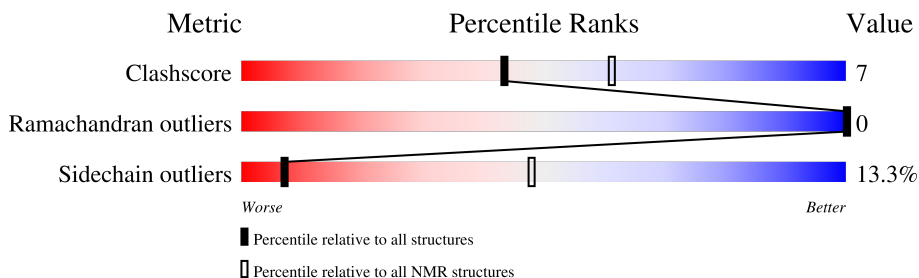
Metric	Whole archive (#Entries)	NMR archive (#Entries)
Clashscore	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428

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	31	74% 10% • 13%

2 Ensemble composition and analysis ⓘ

This entry contains 20 models. Model 9 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *target function*.

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:933-A:948 (16)	0.04	9
2	A:949-A:959 (11)	0.09	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 4 clusters and 3 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 533 atoms, of which 284 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Insulin-like growth factor 1 receptor.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	31	533	172	284	43	33	1	0

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: Insulin-like growth factor 1 receptor

Chain A:
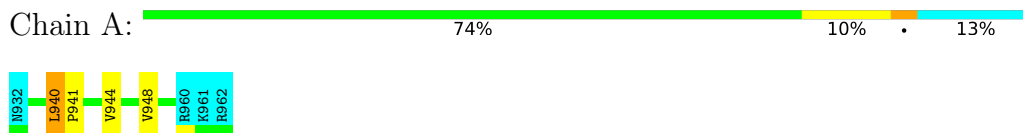
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Insulin-like growth factor 1 receptor

Chain A:

4.2.2 Score per residue for model 2

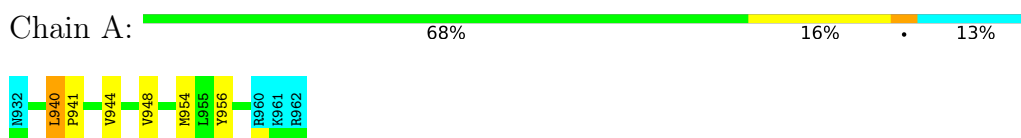• Molecule 1: Insulin-like growth factor 1 receptor

Chain A:

4.2.3 Score per residue for model 3

• Molecule 1: Insulin-like growth factor 1 receptor

Chain A:

4.2.4 Score per residue for model 4

• Molecule 1: Insulin-like growth factor 1 receptor

Chain A:

4.2.5 Score per residue for model 5

• Molecule 1: Insulin-like growth factor 1 receptor

Chain A:

4.2.6 Score per residue for model 6

• Molecule 1: Insulin-like growth factor 1 receptor

Chain A:

4.2.7 Score per residue for model 7

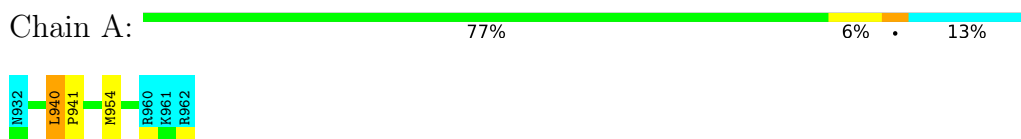• Molecule 1: Insulin-like growth factor 1 receptor

Chain A:

4.2.8 Score per residue for model 8
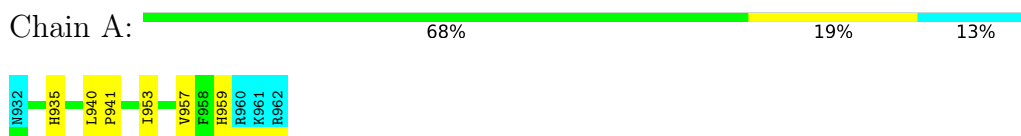
• Molecule 1: Insulin-like growth factor 1 receptor

Chain A:

4.2.9 Score per residue for model 9 (medoid)

• Molecule 1: Insulin-like growth factor 1 receptor

Chain A:

4.2.10 Score per residue for model 10

• Molecule 1: Insulin-like growth factor 1 receptor

Chain A:

4.2.11 Score per residue for model 11

• Molecule 1: Insulin-like growth factor 1 receptor

Chain A:

4.2.12 Score per residue for model 12

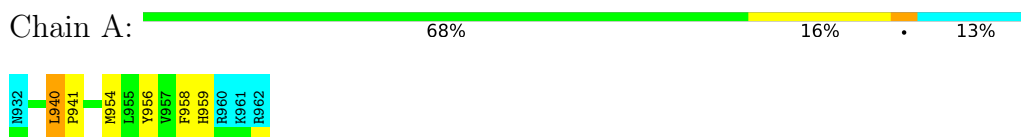• Molecule 1: Insulin-like growth factor 1 receptor

Chain A:

4.2.13 Score per residue for model 13

• Molecule 1: Insulin-like growth factor 1 receptor

Chain A:

4.2.14 Score per residue for model 14

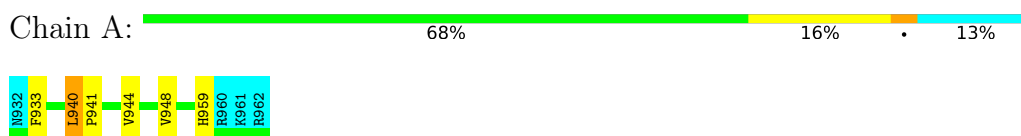• Molecule 1: Insulin-like growth factor 1 receptor

Chain A:

4.2.15 Score per residue for model 15

• Molecule 1: Insulin-like growth factor 1 receptor

Chain A:

4.2.16 Score per residue for model 16

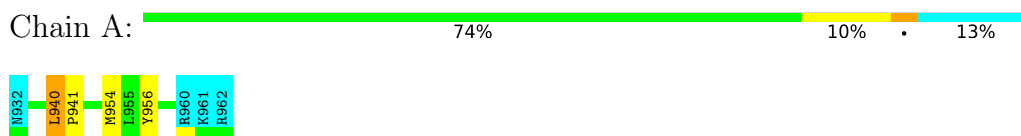• Molecule 1: Insulin-like growth factor 1 receptor

Chain A:

4.2.17 Score per residue for model 17

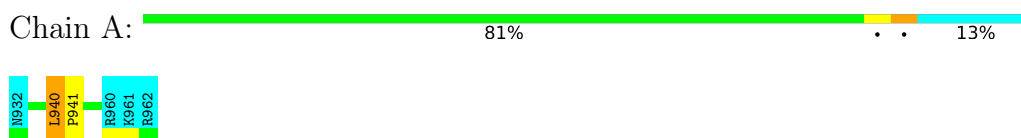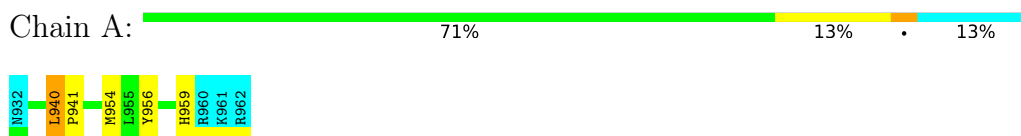• Molecule 1: Insulin-like growth factor 1 receptor

Chain A:

4.2.18 Score per residue for model 18

• Molecule 1: Insulin-like growth factor 1 receptor

Chain A:

77% 6% · 13%

N932 L940 P941 H959 R960 K961 R962
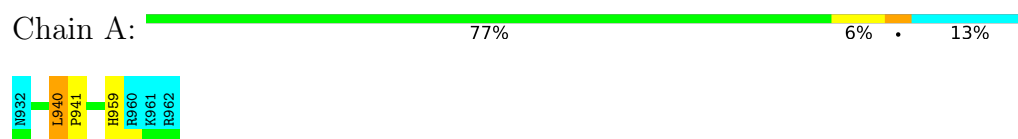
4.2.19 Score per residue for model 19

• Molecule 1: Insulin-like growth factor 1 receptor

Chain A:

68% 16% · 13%

N932 L940 P941 V944 V948 M954 L955 Y956 R960 K961 R962
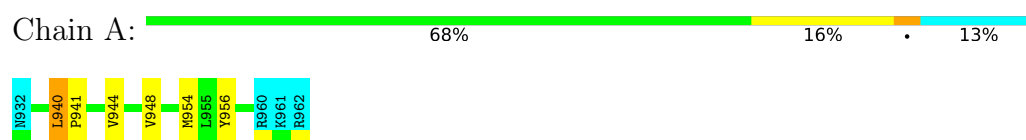
4.2.20 Score per residue for model 20

• Molecule 1: Insulin-like growth factor 1 receptor

Chain A:

65% 23% 13%

N932 I938 A939 L940 P941 V942 A943 V944 V948 Y956 R960 K961 R962
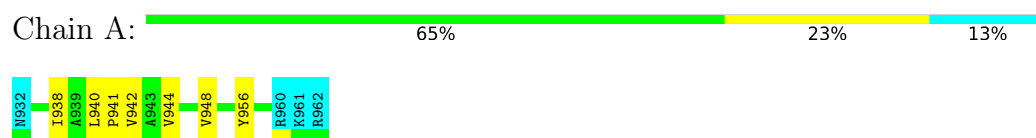
5 Refinement protocol and experimental data overview ⓘ

The models were refined using the following method: *simulated annealing.*

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

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

Software name	Classification	Version
CYANA	refinement	
CYANA	structure calculation	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	454
Number of shifts mapped to atoms	454
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	96%

6 Model quality ⓘ

6.1 Standard geometry ⓘ

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 ⓘ

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	210	239	239	3±1
All	All	4200	4780	4780	63

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 7.

All unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Models Total
1:A:944:VAL:O	1:A:948:VAL:HG23	0.61	1.96	19	11
1:A:948:VAL:O	1:A:952:VAL:HG23	0.59	1.98	16	5
1:A:953:ILE:O	1:A:957:VAL:HG23	0.54	2.03	6	4
1:A:942:VAL:HA	1:A:945:LEU:HD12	0.49	1.85	1	1
1:A:940:LEU:CB	1:A:941:PRO:CD	0.47	2.93	20	20
1:A:940:LEU:N	1:A:941:PRO:HD2	0.43	2.29	12	20
1:A:938:ILE:O	1:A:942:VAL:HG23	0.42	2.15	20	2

6.3 Torsion angles ⓘ

6.3.1 Protein backbone ⓘ

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	27/31 (87%)	27±0 (99±2%)	0±0 (1±2%)	0±0 (0±0%)	100	100
All	All	540/620 (87%)	535 (99%)	5 (1%)	0 (0%)	100	100

There are no Ramachandran outliers.

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	23/27 (85%)	20±1 (87±5%)	3±1 (13±5%)	7	48
All	All	460/540 (85%)	399 (87%)	61 (13%)	7	48

All 7 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	940	LEU	16
1	A	956	TYR	10
1	A	954	MET	10
1	A	959	HIS	8
1	A	958	PHE	7
1	A	933	PHE	5
1	A	935	HIS	5

6.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

6.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.6 Ligand geometry ⓘ

There are no ligands in this entry.

6.7 Other polymers ⓘ

There are no such molecules in this entry.

6.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

7 Chemical shift validation ⓘ

The completeness of assignment taking into account all chemical shift lists is 96% for the well-defined parts and 92% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *IGF1R_structure_dep.str*

7.1.1 Bookkeeping ⓘ

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	454
Number of shifts mapped to atoms	454
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

7.1.2 Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_\alpha$	31	-0.45 ± 0.33	None needed (< 0.5 ppm)
$^{13}C_\beta$	29	0.36 ± 0.08	None needed (< 0.5 ppm)
$^{13}C'$	28	0.46 ± 0.22	None needed (< 0.5 ppm)
^{15}N	30	1.43 ± 0.65	Should be applied

7.1.3 Completeness of resonance assignments ⓘ

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 96%, i.e. 326 atoms were assigned a chemical shift out of a possible 338. 11 out of 11 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	131/133 (98%)	53/53 (100%)	52/54 (96%)	26/26 (100%)
Sidechain	163/165 (99%)	92/93 (99%)	71/72 (99%)	0/0 (—%)

Continued on next page...

Continued from previous page...

	Total	¹H	¹³C	¹⁵N
Aromatic	32/40 (80%)	17/22 (77%)	15/16 (94%)	0/2 (0%)
Overall	326/338 (96%)	162/168 (96%)	138/142 (97%)	26/28 (93%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 92%, i.e. 380 atoms were assigned a chemical shift out of a possible 414. 11 out of 11 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	¹H	¹³C	¹⁵N
Backbone	149/153 (97%)	60/61 (98%)	59/62 (95%)	30/30 (100%)
Sidechain	199/221 (90%)	116/127 (91%)	82/86 (95%)	1/8 (12%)
Aromatic	32/40 (80%)	17/22 (77%)	15/16 (94%)	0/2 (0%)
Overall	380/414 (92%)	193/210 (92%)	156/164 (95%)	31/40 (78%)

7.1.4 Statistically unusual chemical shifts ⓘ

There are no statistically unusual chemical shifts.

7.1.5 Random Coil Index (RCI) plots ⓘ

The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.

Random coil index (RCI) for chain A: