wwPDB NMR Structure Validation Summary Report (i)

Feb 15, 2022 – 08:56 AM EST

PDB ID	:	1KRX
Title	:	SOLUTION STRUCTURE OF BERYLLOFLUORIDE-ACTIVATED NTRC RECEIVER DOMAIN: MODEL STRUCTURES INCORPORATING ACTIVE SITE CONTACTS
Authors	:	Hastings, C.A.; Lee, S.-Y.; Cho, H.S.; Yan, D.; Kustu, S.; Wemmer, D.E.
Deposited on	:	2002-01-10

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:

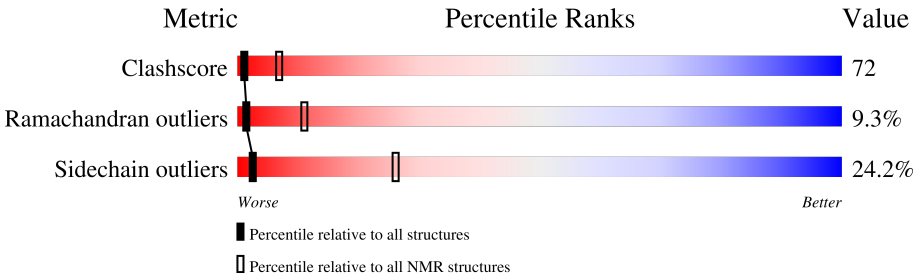
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.26
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal 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	124	16% 65% 15% 5%

2 Ensemble composition and analysis (i)

This entry contains 26 models. Model 6 is the overall representative, medoid model (most similar to other models). The authors have identified model 19 as representative.

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:4-A:121 (118)	0.41	6

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 5 clusters and 9 single-model clusters were found.

Cluster number	Models
1	3, 4, 5, 6, 9, 11, 16, 26
2	12, 18, 19
3	15, 25
4	7, 24
5	21, 22
Single-model clusters	1; 2; 8; 10; 13; 14; 17; 20; 23

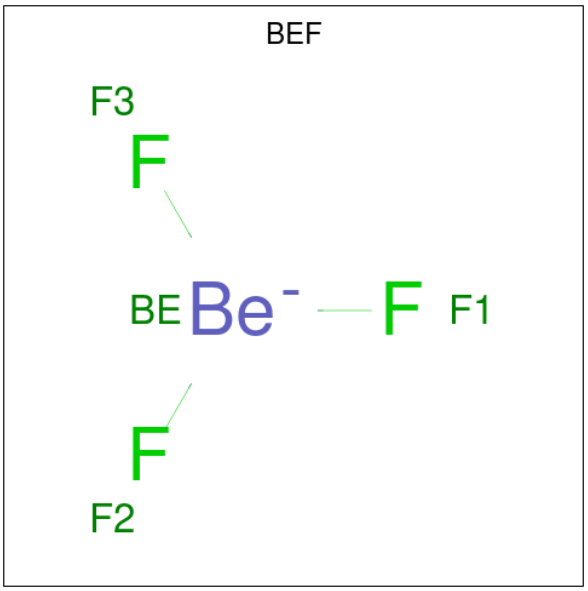
3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1919 atoms, of which 960 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called NITROGEN REGULATION PROTEIN NR(I).

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	124	1915	606	960	162	181	6	0

- Molecule 2 is BERYLLIUM TRIFLUORIDE ION (three-letter code: BEF) (formula: BeF_3).



Mol	Chain	Residues	Atoms		
			Total	Be	F
2	A	1	4	1	3

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: NITROGEN REGULATION PROTEIN NR(I)

Chain A: 16% 65% 15% 5%

M1 Q2 R3 G4 I5 V6 W7 V8 V9 D10 D11 D12 S13 S14 I15 R16 W17 V18 L19 E20 R21 A22 L23 A26 G27 L28 T29 C30 T31 T32 F33 E34 N35 G36 N37 E38 V39 L40 L43 A44 S45 K46 T47 P48 D49 V50 L51 L52 S53 D54 I55 R56 M57 P58 G59 M60 D61 G62

L63 A64 L65 L66 K67 Q68 I69 K70 Q71 R72 H73 P74 M75 L76 P77 V78 I79 I80 M81 T82 A83 H84 S85 D86 L87 D88 A89 A90 V91 S92 A93 Y94 Q95 Q96 G97 A98 F99 D100 Y101 L102 P103 K104 P105 F106 D107 I108 D109 E110 A111 V112 A113 L114 V115 E116 R117 A118 Y122 Q123 E124

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

The representative model is number 6. Colouring as in section 4.1 above.

• Molecule 1: NITROGEN REGULATION PROTEIN NR(I)

Chain A: 18% 60% 15% • 5%

M1 Q2 R3 G4 I5 V6 W7 V8 V9 D10 D11 S14 I15 R16 W17 V18 L19 E20 R21 A22 L23 A26 G27 L28 T29 C30 T31 T32 F33 E34 N35 G36 N37 E38 V39 L40 L43 A44 S45 K46 T47 P48 D49 V50 L51 L52 S53 D54 I55 R56 M57 P58 G59 M60 D61 G62 L63

A64 L65 L66 K67 Q68 I69 K70 Q71 R72 H73 P74 M75 L76 P77 V78 I79 I80 M81 T82 A83 H84 S85 D86 L87 D88 A89 A90 V91 S92 A93 Y94 Q95 Q96 G97 A98 F99 D100 Y101 L102 P103 K104 P105 F106 D107 I108 D109 E110 A111 V112 A113 L114 V115 E116 R117 Y122 Q123 E124

5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 26 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
DYANA	structure solution	1.5
CNS 1.0	structure solution	
CNS 1.0	refinement	
MOLMOL	refinement	

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: BEF

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	897	905	902	130±12
All	All	23426	23530	23452	3393

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 72.

5 of 870 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:23:LEU:HD22	1:A:28:LEU:HD13	1.11	1.22	22	3
1:A:9:VAL:HG13	1:A:33:PHE:HB2	1.06	1.26	16	23
1:A:6:VAL:HG11	1:A:23:LEU:HD13	1.02	1.22	21	4
1:A:66:LEU:HD13	1:A:93:ALA:HB2	0.99	1.32	4	4
1:A:7:TRP:CD1	1:A:31:THR:HG22	0.98	1.94	23	26

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	118/124 (95%)	92±3 (78±3%)	15±3 (13±3%)	11±2 (9±1%)	1 11
All	All	3068/3224 (95%)	2397 (78%)	387 (13%)	284 (9%)	1 11

5 of 25 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	106	PHE	26
1	A	107	ASP	26
1	A	55	ILE	21
1	A	100	ASP	20
1	A	13	SER	19

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	96/102 (94%)	73±3 (76±3%)	23±3 (24±3%)	2 26
All	All	2496/2652 (94%)	1893 (76%)	603 (24%)	2 26

5 of 60 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	LEU	26
1	A	104	LYS	26
1	A	81	MET	25
1	A	8	VAL	24
1	A	46	LYS	24

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)

1 ligand 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	BEF	A	125	1	0,3,3	0.00±0.00	-

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	BEF	A	125	1	-	-	-

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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